and

$$\mathbf{i}\times\mathbf{j}=\mathbf{k},\quad \mathbf{j}\times\mathbf{i}=-\mathbf{k},\quad \mathbf{j}\times\mathbf{k}=\mathbf{i},\quad \mathbf{k}\times\mathbf{j}=-\mathbf{i},\quad \mathbf{k}\times\mathbf{i}=\mathbf{j},\ \ \mathbf{i}\times\mathbf{k}=-\mathbf{j}. \tag{41}$$

Only results (35) and (36) require some comment, as the other results are obvious. The change of sign in (35) that makes the cross product *anticommutative* occurs because when the vectors **a** and **b** are interchanged, the right-hand rule causes the direction of $\hat{\mathbf{n}}$ to be reversed. Result (36) can be proved in several ways, but we shall postpone its proof until a different expression for the cross product has been derived.

To obtain a more convenient expression for the cross product that can be used when **a** and **b** are known in terms of their components, we proceed as follows. Let $\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$ and $\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}$, and consider the cross product $\mathbf{a}\times\mathbf{b}=(a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k})\times(b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k})$. Expanding this expression term by term is justified because of the associative property given in (36), and it leads to the result

$$\begin{aligned}\mathbf{a}\times\mathbf{b}&=a_1b_1\mathbf{i}\times\mathbf{i}+a_1b_2\mathbf{i}\times\mathbf{j}+a_1b_3\mathbf{i}\times\mathbf{k}+a_2b_1\mathbf{j}\times\mathbf{i}+a_2b_2\mathbf{j}\times\mathbf{j}\\&\quad+a_2b_3\mathbf{j}\times\mathbf{k}+a_3b_1\mathbf{k}\times\mathbf{i}+a_3b_2\mathbf{k}\times\mathbf{j}+a_3b_3\mathbf{k}\times\mathbf{k}.\end{aligned}$$

cross product in terms of components

Results (40) cause three terms on the right-hand side to vanish, and results (41) allow the remaining six terms to be collected into three groups as follows to give

$$\mathbf{a}\times\mathbf{b}=(a_2b_3-a_3b_2)\mathbf{i}-(a_1b_3-a_3b_1)\mathbf{j}+(a_1b_2-a_2b_1)\mathbf{k}. \tag{42}$$

This alternative expression for the cross product in terms of the cartesian components of vectors **a** and **b** can be further simplified by making formal use of the third-order determinant,

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\a_1&a_2&a_3\\b_1&b_2&b_3\end{vmatrix},$$

because a formal expansion in terms of elements of the first row generates result (42). We take this result as an alternative but equivalent definition of the cross product.

practical definition of a cross product using a determinant

The cross product (cartesian component form)

Let $\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$ and $\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}$. Then

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\a_1&a_2&a_3\\b_1&b_2&b_3\end{vmatrix}, \tag{43}$$

When expressing $\mathbf{a}\times\mathbf{b}$ as the determinant in (43), purely *formal* use was made of the method of expansion of a determinant in terms of the elements of its first row, because (43) is not a determinant in the ordinary sense as its elements are a mixture of vectors and numbers.

EXAMPLE 2.16 Given that $\mathbf{a} = 3\mathbf{i} - 2\mathbf{j} - \mathbf{k}$ and $\mathbf{b} = \mathbf{i} + 4\mathbf{j} + 2\mathbf{k}$, find $\mathbf{a} \times \mathbf{b}$ and a unit vector $\hat{\mathbf{n}}$ normal to the plane containing $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{n}$, in this order, obey the right-hand rule.

Solution Substitution into expression (43) gives

$$\begin{aligned}\mathbf{a} \times \mathbf{b} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 3 & -2 & -1 \\ 1 & 4 & 2 \end{vmatrix} \\ &= [(-2)\cdot 2 - 4\cdot(-1)]\mathbf{i} - [3\cdot 2 - 1\cdot(-1)]\mathbf{j} + [3\cdot 4 - 1\cdot(-2)]\mathbf{k} \\ &= -7\mathbf{j} + 14\mathbf{k}.\end{aligned}$$

The required unit vector $\hat{\mathbf{n}}$ is simply the unit vector in the direction of $\mathbf{a} \times \mathbf{b}$, so

$$\begin{aligned}\hat{\mathbf{n}} &= (\mathbf{a} \times \mathbf{b})/\|\mathbf{a} \times \mathbf{b}\| = (-7\mathbf{j} + 14\mathbf{k})/(7\sqrt{5}). \\ &= (-1/\sqrt{5})\mathbf{j} + (2/\sqrt{5})\mathbf{k}.\end{aligned}$$

■

We now return to the proof of the associative property stated in (35) and establish it by means of result (43).

Setting $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, and $\mathbf{c} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$, we have

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ (b_1 + c_1) & (b_2 + c_2) & (b_3 + c_3) \end{vmatrix}.$$

Expanding the determinant in terms of elements of its first row and grouping terms gives

$$\begin{aligned}\mathbf{a} \times (\mathbf{b} + \mathbf{c}) &= (a_2b_3 - a_3b_2)\mathbf{i} - (a_1b_3 - a_3b_1)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \\ &\quad + (a_2c_3 - a_3c_2)\mathbf{i} - (a_1c_3 - a_3c_1)\mathbf{j} + (a_1c_2 - a_2c_1)\mathbf{k} \\ &= \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c},\end{aligned}$$

and the result is proved.

Summary

This section first introduced the vector or cross product of two vectors in geometrical terms and then used the result to show that the vector product is anticommutative, in the sense that $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$. Important results involving the vector product are given in terms of the components of the two vectors that are involved. Finally, the vector product was expressed in a form that is most convenient for calculations by writing it in determinantal form, the rows of which contain the unit vectors **i**, **j**, and **k** and the components of the respective vectors.

EXERCISES 2.3

In Exercises 1 through 6 use (43) to find $\mathbf{a} \times \mathbf{b}$.

1. For $\mathbf{a} = 2\mathbf{i} - \mathbf{j} - 4\mathbf{k}$, $\mathbf{b} = 3\mathbf{i} - \mathbf{j} - \mathbf{k}$.

2. For $\mathbf{a} = -3\mathbf{i} + 2\mathbf{j} + 4\mathbf{k}$, $\mathbf{b} = 2\mathbf{i} + \mathbf{j} - 2\mathbf{k}$.

3. For $\mathbf{a} = 7\mathbf{i} + 6\mathbf{k}$, $\mathbf{b} = 3\mathbf{j} + \mathbf{k}$.

4. For $\mathbf{a} = 3\mathbf{i} + 7\mathbf{j} + 2\mathbf{k}$, $\mathbf{b} = \mathbf{i} - \mathbf{j} + \mathbf{k}$.

5. For $\mathbf{a} = 2\mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} - \mathbf{j} + \mathbf{k}$.

6. For $\mathbf{a} = 3\mathbf{i} - 2\mathbf{j} + 6\mathbf{k}$, $\mathbf{b} = 2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$.

In Exercises 7 through 10 verify the equivalence of the definitions of the cross product in (34) and (43) by first using (43) to calculate $\mathbf{a} \times \mathbf{b}$, and hence $\|\mathbf{a} \times \mathbf{b}\|$ and $\hat{\mathbf{n}}$, and then calculating $\|\mathbf{a}\|$ and $\|\mathbf{b}\|$ directly, using result (17) to find $\cos\theta$ and hence $\sin\theta$, and using the results to find $\mathbf{a} \times \mathbf{b}$ from (34).

7. For $\mathbf{a} = \mathbf{i} + \mathbf{j} + 3\mathbf{k}$ and $\mathbf{b} = 3\mathbf{i} + 2\mathbf{j} + \mathbf{k}$.

8. For $\mathbf{a} = \mathbf{i} + \mathbf{j} + \mathbf{k}$ and $\mathbf{b} = 4\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$.

9. For $\mathbf{a} = 2\mathbf{i} + \mathbf{j} - 3\mathbf{k}$ and $\mathbf{b} = 5\mathbf{i} - 2\mathbf{k}$.

10. For $\mathbf{a} = -2\mathbf{i} - 3\mathbf{j} + \mathbf{k}$ and $\mathbf{b} = 3\mathbf{i} + \mathbf{j} + 2\mathbf{k}$.

In Exercises 11 through 14, verify by direct calculation that $(\mathbf{b} + \mathbf{c}) \times \mathbf{a} = -\mathbf{a} \times (\mathbf{b} + \mathbf{c})$.

11. $\mathbf{a} = 3\mathbf{j} + 2\mathbf{k}$, $\mathbf{b} = \mathbf{i} - 4\mathbf{j} + \mathbf{k}$, and $\mathbf{c} = 5\mathbf{i} - 2\mathbf{j} + 3\mathbf{k}$.

12. $\mathbf{a} = -\mathbf{i} + 5\mathbf{j} + 2\mathbf{k}$, $\mathbf{b} = 4\mathbf{i} + \mathbf{k}$, and $\mathbf{c} = -2\mathbf{i} - 4\mathbf{j} + 3\mathbf{k}$.

13. $\mathbf{a} = \mathbf{i} + \mathbf{k}$, $\mathbf{b} = 3\mathbf{i} - \mathbf{j} - 2\mathbf{k}$, and $\mathbf{c} = 3\mathbf{i} + \mathbf{j} + \mathbf{k}$.

14. $\mathbf{a} = 5\mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} - \mathbf{j} - \mathbf{k}$, and $\mathbf{c} = 4\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.

In Exercises 15 through 18 find a unit vector normal to a plane containing the given vectors.

15. $3\mathbf{i} + \mathbf{j} + \mathbf{k}$ and $\mathbf{i} + 2\mathbf{j} + \mathbf{k}$.

16. $2\mathbf{i} - \mathbf{j} + 2\mathbf{k}$ and $2\mathbf{i} + 3\mathbf{j} + \mathbf{k}$.

17. $\mathbf{i} + \mathbf{j} + \mathbf{k}$ and $2\mathbf{i} + 3\mathbf{j} - \mathbf{k}$.

18. $2\mathbf{i} + 2\mathbf{j} - \mathbf{k}$ and $3\mathbf{i} + \mathbf{j} + 4\mathbf{k}$.

19. Find a unit vector normal to a plane containing vectors $\mathbf{a} + \mathbf{b}$ and $\mathbf{a} + \mathbf{c}$, given that $\mathbf{a} = \mathbf{i} + 2\mathbf{j} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} + \mathbf{j} - 2\mathbf{k}$, and $\mathbf{c} = 3\mathbf{i} + 2\mathbf{j} + 4\mathbf{k}$.

20. Given that $\mathbf{a} = 3\mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} - \mathbf{j} + 2\mathbf{k}$, and $\mathbf{c} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, find (a) a vector normal to the plane containing the vectors $\mathbf{a} + (\mathbf{a} \cdot \mathbf{b})\mathbf{b}$ and $\mathbf{c}$ and, (b) explain why the normal to a plane containing the vectors $\mathbf{a}$ and $\mathbf{b}$ and the normal to a plane containing the vectors $(\mathbf{a} \cdot \mathbf{b})\mathbf{a}$ and $(\mathbf{b} \cdot \mathbf{c})\mathbf{b}$ are parallel.

In Exercises 21 through 24, find the cartesian equation of the plane that passes through the given points.

21. $(1, 3, 2)$, $(2, 0, -4)$, and $(1, 6, 11)$.

22. $(1, 4, 3)$, $(2, 0, 1)$, and $(3, 4, -6)$.

23. $(1, 2, 3)$, $(2, -4, 1)$, and $(3, 6, -1)$.

24. $(1, 0, 1)$, $(2, 5, 7)$, and $(2, 3, 9)$.

Three points with position vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ will be **collinear** (lie on a line) if the parallelogram with adjacent sides $\mathbf{a} - \mathbf{b}$ and $\mathbf{a} - \mathbf{c}$ has zero geometrical area. Use this result in Exercises 25 through 28 to determine which sets of points are collinear.

25. $(2, 2, 3)$, $(6, 1, 5)$, $(-2, 4, 3)$.

26. $(1, 2, 4)$, $(7, 0, 8)$, $(-8, 5, -2)$.

27. $(2, 3, 3)$, $(3, 7, 5)$, $(0, -5, -1)$.

28. $(1, 3, 2)$, $(4, 2, 1)$, $(1, 0, 2)$.

29. A vector $\mathbf{N}$ normal to the plane containing the skew vectors $\mathbf{a}$ and $\mathbf{b}$ can be found as follows. $\mathbf{N}$ is normal to $\mathbf{a}$ and $\mathbf{b}$, so $\mathbf{a} \cdot \mathbf{N} = 0$ and $\mathbf{b} \cdot \mathbf{N} = 0$. If a component of $\mathbf{N}$ is assigned an arbitrary nonzero value c, say, the other two components can be found from these two equations as multiples of c, and $\mathbf{N}$ will then be determined as a multiple of c. A suitable choice of c will make $\mathbf{N}$ a unit normal $\hat{\mathbf{N}}$. Apply this method to vectors $\mathbf{a}$ and $\mathbf{b}$ in Exercise 7 to find a vector $\hat{\mathbf{N}}$. Compare the result with the unit vector

$$\hat{\mathbf{n}} = (\mathbf{a} \times \mathbf{b})/\|\mathbf{a} \times \mathbf{b}\|$$

found from (43). Explain why although both $\hat{\mathbf{n}}$ and $\hat{\mathbf{N}}$ are normal to the plane containing $\mathbf{a}$ and $\mathbf{b}$ they may have opposite senses.

2.4 Linear Dependence and Independence of Vectors and Triple Products

The dot and cross products can be combined to provide a simple test that determines whether or not an arbitrary set of three vectors possesses a property of fundamental importance to the algebra of vectors. First, however, some introductory remarks are necessary.

Given a set of n vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$, and a set of n constants $c_1, c_2, \ldots, c_n$, the sum

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$$

linear combination of vectors

is called a **linear combination** of the vectors. Linear combinations of the vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ were used in Section 2.1 to express *every* vector in three-dimensional space as a linear combination of these three vectors. A triad of vectors such as $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ with the property that *all* vectors in three-dimensional space can be represented as linear combinations of these three vectors is said to form a **basis** for the space.

basis

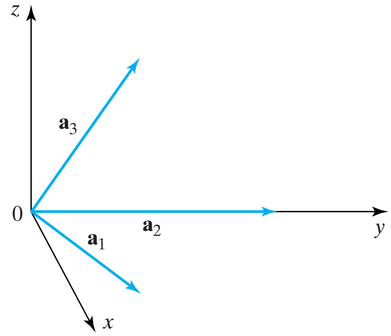


FIGURE 2.20 Nonorthogonal triad forming a basis in three-dimensional space.

It is a fundamental property of three-dimensional space that a basis for the space comprises a set of three vectors $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_3$, with the property that the linear combination

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3 = \mathbf{0} \tag{44}$$

is *only* true when $c_1 = c_2 = c_3 = 0$. Vectors $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_3$ satisfying this condition are said to be **linearly independent** vectors, and a vector $\mathbf{d}$ of the form

linear independence and linear dependence

$$\mathbf{d} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3,$$

where *not all* of c_1, c_2, and c_3 are zero, is said to be **linearly dependent** on the vectors $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_3$. The vectors $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ that form a basis for three-dimensional space are linearly independent vectors, but the position vector $\mathbf{r} = 2\mathbf{i} - 3\mathbf{j} + 5\mathbf{k}$ is linearly dependent on vectors $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$.

Clearly, vectors $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ do not form the only basis for three-dimensional space, because any triad of linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_3$ will serve equally well, as, for example, the nonorthogonal set of vectors shown in Fig. 2.20.

The dot and cross products will now be combined to develop a test for linear dependence and independence based on the elementary geometrical idea of the volume of the parallelepiped shown in Fig. 2.21, three edges $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ of which meet at the origin.

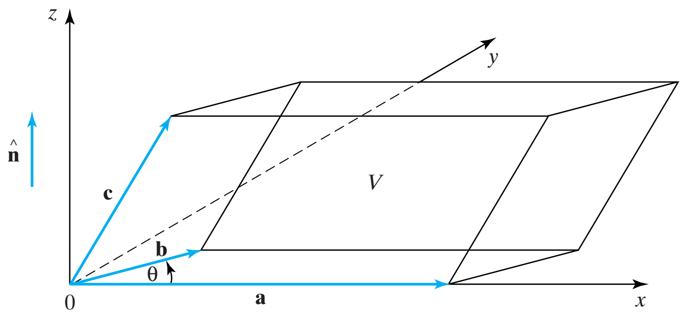


FIGURE 2.21 Volume V of a parallelepiped.

The volume V of a parallelepiped is a nonnegative number given by the product of the area of its base and its height. Suppose vectors $\mathbf{a}$ and $\mathbf{b}$ are chosen to form two sides of the base of the parallelepiped. Then the vector area of this base has already been interpreted as $\mathbf{a} \times \mathbf{b}$. The vertical height of the parallelepiped is the projection of vector $\mathbf{c}$ in the direction of the unit vector $\hat{\mathbf{n}}$ normal to the base, and so is given by $\hat{\mathbf{n}} \cdot \mathbf{c}$. Consequently, as $\mathbf{a} \times \mathbf{b} = \|\mathbf{a}\| \cdot \|\mathbf{b}\| \sin\theta\hat{\mathbf{n}}$, it follows that

$$V = |(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|. \tag{45}$$

The absolute value of the right-hand side of (45) has been taken because a volume must be a nonnegative quantity, whereas the dot product $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ may be of either sign.

If vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ form a basis for three-dimensional space, vector $\mathbf{c}$ cannot be linearly dependent on vectors $\mathbf{a}$ and $\mathbf{b}$, and so the parallelepiped in Fig. 2.21 with these vectors as its sides must have a nonzero volume. If, however, vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are **coplanar** (all lie in the same plane), and so cannot form a basis for the space, the volume of the parallelepiped will be zero. These simple geometrical observations lead to the following test for the linear independence of three vectors in three-dimensional space.

THEOREM 2.3

a test for linear independence

Test for linear independence of vectors in three-dimensional space Let $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ be any three vectors. Then the vectors are linearly independent if $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} \neq 0$, and they are linearly dependent if $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = 0$. ■

A product of the type $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ is called a **scalar triple product**. The name arises because the result is a scalar. It is also called a mixed triple product since both $\cdot$ and $\times$ appear. Three vectors are involved in this dot (scalar) product, one of which is the vector $\mathbf{a} \times \mathbf{b}$ and the other is the vector $\mathbf{c}$.

scalar triple product

Scalar triple products are easily evaluated, because taking the dot product of $\mathbf{a} \times \mathbf{b}$ in the form given in (42) with $\mathbf{c} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$ gives

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = (a_2b_3 - a_3b_2)c_1 - (a_1b_3 - a_3b_1)c_2 + (a_1b_2 - a_2b_1)c_3.$$

The right-hand side of this expression is simply the value of a determinant with successive rows given by the components of $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, so we have arrived at the following convenient formula for the scalar triple product.

scalar triple product as a determinant

Scalar triple product

Let $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, and $\mathbf{c} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$. Then

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \tag{46}$$

Interchanging any two rows in a matrix changes the sign but not the value of its determinant. Two such switches in (46) leave the value unchanged, so the dot

product is commutative and so we arrive at the useful result

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}). \tag{47}$$

So, in a scalar triple product the dot and cross may be *interchanged* without altering the result.

EXAMPLE 2.17

Given the two sets of vectors (a) $\mathbf{a} = \mathbf{i} + 2\mathbf{j} - 5\mathbf{k}$, $\mathbf{b} = \mathbf{i} + \mathbf{j} + 2\mathbf{k}$, $\mathbf{c} = \mathbf{i} + 4\mathbf{j} - 19\mathbf{k}$ and (b) $\mathbf{a} = 2\mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{b} = 3\mathbf{i} + 4\mathbf{k}$, $\mathbf{c} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, find if the vectors are linearly independent or linearly dependent.

Solution We apply Theorem 2.3 to each set, using result (46) to evaluate the scalar triple products.

(a)
$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \begin{vmatrix} 1 & 2 & -5 \\ 1 & 1 & 2 \\ 1 & 4 & -19 \end{vmatrix} = 0,$$

so the set of three vectors in (a) is linearly dependent. In fact this can be seen from the fact that $\mathbf{c} = 3\mathbf{a} - 2\mathbf{b}$.

(b)
$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \begin{vmatrix} 2 & 1 & 1 \\ 3 & 0 & 4 \\ 1 & 1 & 1 \end{vmatrix} = -4 \neq 0,$$

so the set of three vectors in (b) is linearly independent. Although not required, the volume V of the parallelepiped formed by these three vectors is $V = |(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}| = |-4| = 4$. ■

Another notation for the scalar triple product of vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ is $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$, so

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}, \tag{48}$$

or, in terms of a determinant,

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \tag{49}$$

alternative forms of a scalar triple product

Using this definition of $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ with the row interchange property of determinants (see Section 1.7) shows that

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = [\mathbf{b}, \mathbf{c}, \mathbf{a}] = [\mathbf{c}, \mathbf{a}, \mathbf{b}], \tag{50}$$

because two row interchanges are needed to arrive at $[\mathbf{b}, \mathbf{c}, \mathbf{a}]$ from $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$, leaving the sign of the determinant unchanged, whereas two more are required to arrive at $[\mathbf{c}, \mathbf{a}, \mathbf{b}]$ from $[\mathbf{b}, \mathbf{c}, \mathbf{a}]$, again leaving the sign of the determinant unchanged.

The order of the vectors in results (46), or in the equivalent notation of (48), is easily remembered when the results are abbreviated to

$$\begin{matrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \\ \mathbf{b} & \mathbf{c} & \mathbf{a} \\ \mathbf{c} & \mathbf{a} & \mathbf{b} \end{matrix}$$

In this pattern, row two follows from row one when the first letter is moved to the end position, and row three follows from row two by means of the same process. The effect of applying this process to the third row is simply to regenerate the first row. Rearrangements of this kind are called **cyclic permutations** of the three vectors.

Again making use of the row interchange property of determinants (see Section 1.7), it follows that

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = -[\mathbf{a}, \mathbf{c}, \mathbf{b}],$$

because this time only one row interchange is needed to produce the result on the right from the one on the left, so that a sign change is involved.

A different product involving the three vectors **a**, **b**, and **c** that this time generates another vector is of the form

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}),$$

and products of this type are called **vector triple products** since the results are vectors. In these products it is essential to include the brackets because, in general, $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$. The most important results concerning vector triple products are given in the following theorem.

THEOREM 2.4

Vector triple products If **a**, **b**, and **c** are any three vectors, then

(a)
$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$$

vector triple product

and

(b)
$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}.$$

Proof The proof of the results in Theorem 2.4 both follow in similar fashion, so we only prove result (a) and leave the proof of result (b) as an exercise. We write the cross product $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ in the form of the determinant in (43), with the components of **a** in the second row and those of $\mathbf{b} \times \mathbf{c}$ (obtained from (42)) in the third row when we find that

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ (b_2c_3 - b_3c_2) & (b_3c_1 - b_1c_3) & (b_1c_2 - b_2c_1) \end{vmatrix}.$$

Expanding this determinant in terms of the elements of its first row and grouping terms gives

$$\begin{aligned}\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = {} & [(a_2c_2 + a_3c_3)b_1 - (a_2b_2 + a_3b_3)c_1]\mathbf{i} + [(a_1c_1 + a_3c_3)b_2 \\ & - (a_1b_1 + a_3b_3)c_2]\mathbf{j} + [(a_1c_1 + a_2c_2)b_3 - (a_1b_1 + a_2b_2)c_3]\mathbf{k}.\end{aligned}$$

As it stands, this result is not yet in the form that is required, but adding and subtracting $a_1b_1c_1$ to the coefficient of **i**, $a_2b_2c_2$ to the coefficient of **j**, and $a_3b_3c_3$ to the coefficient of **k** followed by grouping terms give

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c},$$

and the result is established. ■

EXAMPLE 2.18

Find $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ and $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$, given that $\mathbf{a} = 3\mathbf{i} + \mathbf{j} - 4\mathbf{k}$, $\mathbf{b} = 2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$, and $\mathbf{c} = \mathbf{i} + 5\mathbf{j} - \mathbf{k}$.

Solution $\mathbf{a} \cdot \mathbf{b} = -5$, $\mathbf{a} \cdot \mathbf{c} = 12$, and $\mathbf{b} \cdot \mathbf{c} = 4$, so

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c} = 12\mathbf{b} + 5\mathbf{c} = 29\mathbf{i} + 37\mathbf{j} + 31\mathbf{k},$$

and

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a} = 12\mathbf{b} - 4\mathbf{a} = 12\mathbf{i} + 8\mathbf{j} + 52\mathbf{k}. \quad \blacksquare$$

Accounts of geometrical vectors can be found, for example, in references [2.1], [2.3], [2.6], and [1.6].

Summary

This section introduced the two fundamental concepts of linear dependence and independence of vectors. It then showed how the scalar triple product involving three vectors, that gives rise to a scalar quantity, provides a simple test for the linear dependence or independence of the vectors involved. A simple and convenient way of calculating a scalar triple product was shown to be in terms of a determinant with the elements in its rows formed by the components of the three vectors involved in the product. Finally a vector triple product was defined that gives rise to a vector quantity, and it was shown that to avoid ambiguity it is necessary to bracket a pair of vectors in such a product. A rule for the expansion of a vector triple product was derived and shown to involve a linear combination of two of the vectors multiplied by scalar products so that, for example, $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$.

EXERCISES 2.4

In Exercises 1 through 4 use the vectors **a**, **b**, and **c** to find (a) the scalar triple product $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$, and (b) the volume of the parallelepiped determined by these three vectors directed away from a corner.

1. $\mathbf{a} = 2\mathbf{i} - \mathbf{j} - 3\mathbf{k}$, $\mathbf{b} = 3\mathbf{i} - 2\mathbf{k}$, $\mathbf{c} = \mathbf{i} + \mathbf{j} - 4\mathbf{k}$.

2. $\mathbf{a} = \mathbf{i} - \mathbf{j} + 2\mathbf{k}$, $\mathbf{b} = \mathbf{i} + \mathbf{j} + 3\mathbf{k}$, $\mathbf{c} = 2\mathbf{i} - \mathbf{j} + 3\mathbf{k}$.

3. $\mathbf{a} = -\mathbf{i} - \mathbf{j} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$, $\mathbf{c} = -4\mathbf{i} + \mathbf{j} + 3\mathbf{k}$.

4. $\mathbf{a} = 5\mathbf{i} + 3\mathbf{k}$, $\mathbf{b} = 2\mathbf{i} - \mathbf{j}$, $\mathbf{c} = -2\mathbf{i} + 3\mathbf{j} - 2\mathbf{k}$.

In Exercises 5 through 10 find which sets of vectors are coplanar.

5. $\mathbf{i} + 3\mathbf{j} + 2\mathbf{k}$, $2\mathbf{i} + \mathbf{j} + 4\mathbf{k}$, $4\mathbf{i} + 7\mathbf{j} + 8\mathbf{k}$.

6. $2\mathbf{i} + \mathbf{j} + 4\mathbf{k}$, $\mathbf{i} + 2\mathbf{j} + \mathbf{k}$, $4\mathbf{i} + 3\mathbf{j} + 6\mathbf{k}$.

7. $2\mathbf{i} + \mathbf{k}$, $\mathbf{i} + 4\mathbf{j} + 2\mathbf{k}$, $3\mathbf{i} + 12\mathbf{j} + 7\mathbf{k}$.

8. $\mathbf{i} + \mathbf{j} + \mathbf{k}$, $2\mathbf{i} + \mathbf{j} + 2\mathbf{k}$, $4\mathbf{i} + 3\mathbf{j} + \mathbf{k}$.

9. $2\mathbf{i} + \mathbf{j} - \mathbf{k}$, $3\mathbf{i} + \mathbf{j} + 2\mathbf{k}$, $5\mathbf{i} + \mathbf{j} + 8\mathbf{k}$.

10. $2\mathbf{i} + \mathbf{j} - \mathbf{k}$, $\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$, $5\mathbf{i} + 4\mathbf{j} + \mathbf{k}$.

In Exercises 11 through 15 use computer algebra to verify that $[\mathbf{a}, \mathbf{b}, \mathbf{c}] = [\mathbf{c}, \mathbf{a}, \mathbf{b}] = -[\mathbf{a}, \mathbf{c}, \mathbf{b}]$.

11. $\mathbf{a} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} + \mathbf{j} - \mathbf{k}$, $\mathbf{c} = 3\mathbf{i} - \mathbf{j} + \mathbf{k}$.

12. $\mathbf{a} = \mathbf{i} - \mathbf{j} - \mathbf{k}$, $\mathbf{b} = -5\mathbf{i} + 2\mathbf{j} - 3\mathbf{k}$, $\mathbf{c} = 2\mathbf{i} + 3\mathbf{j} - 2\mathbf{k}$.

13. $\mathbf{a} = -3\mathbf{i} - 4\mathbf{j} + \mathbf{k}$, $\mathbf{b} = 9\mathbf{i} + 12\mathbf{j} - 3\mathbf{k}$, $\mathbf{c} = \mathbf{i} + 2\mathbf{j} + \mathbf{k}$.

14. $\mathbf{a} = 3\mathbf{i} + 4\mathbf{k}$, $\mathbf{b} = \mathbf{i} + 5\mathbf{k}$, $\mathbf{c} = 2\mathbf{j} + \mathbf{k}$.

15. Prove that if **a**, **b**, **c**, and **d** are any four vectors, and λ, μ are arbitrary scalars $[\lambda\mathbf{a} + \mu\mathbf{b}, \mathbf{c}, \mathbf{d}] = \lambda[\mathbf{a}, \mathbf{c}, \mathbf{d}] + \mu[\mathbf{b}, \mathbf{c}, \mathbf{d}]$. Use computer algebra with vectors **a**, **b**, **c**, **d** from Exercise 12 with $\mathbf{d} = 4\mathbf{c} - 2\mathbf{j} + 6\mathbf{k}$, and scalars λ, μ of your choice, to verify this result.

In Exercises 16 through 20 find (a) the cartesian equation of the plane containing the given points, and (b) a unit vector normal to the plane.

16. $(1, 2, 1)$, $(3, 1, -2)$, $(2, 1, 4)$.

17. $(2, 0, 3)$, $(0, 1, 0)$, $(2, 4, 5)$.

18. $(-1, 2, -3)$, $(2, 4, 1)$, $(3, 0, 1)$.

19. $(1, 2, 5)$, $(-2, 1, 0)$, $(0, 2, 0)$.

20. Prove result (b) of Theorem 2.4.

21. Show that

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) + \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = \mathbf{0}.$$

22. The **law of sines** for a triangle with angles A, B, and C opposite sides with the respective lengths a, b, and c

takes the form

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}.$$

Prove this by considering a vector triangle with sides **a**, **b**, and **c**, where $\mathbf{c} = \mathbf{a} + \mathbf{b}$, and taking the cross product of $\mathbf{c} = \mathbf{a} + \mathbf{b}$ first with **a**, then with **b**, and finally with **c**.

In Exercises 23 through 26 use the fact that four points with position vectors **p**, **q**, **r**, and **s** will be coplanar if the vectors $\mathbf{p} - \mathbf{q}$, $\mathbf{p} - \mathbf{r}$, and $\mathbf{p} - \mathbf{s}$ are coplanar to find which sets of points all lie in a plane.

23. $(1, 1, -1), (-3, 1, 1), (-1, 2, -1), (1, 0, 0)$.

24. $(1, 2, -1), (2, 1, 1), (0, 1, 2), (1, 1, 1)$.

25. $(0, -4, 0), (2, 3, 1), (3, -4, -2), (4, -2, -2)$.

26. $(1, 2, 3), (1, 0, 1), (2, 1, 2), (4, 1, 0)$.

27. The volume of a tetrahedron is one-third of the product of the area of its base and its vertical height. Show the volume V of the tetrahedron in Fig. 2.22, in which three edges formed by the vectors **a**, **b**, and **c** are directed away from a vertex, is given by

$$V = (1/6)|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$$

28. Let **a**, **b**, **c**, and **d** be vectors and λ, μ, ν be scalars satisfying the equation

$$\lambda(\mathbf{b} \times \mathbf{c}) + \mu(\mathbf{c} \times \mathbf{a}) + \nu(\mathbf{a} \times \mathbf{b}) + \mathbf{d} = \mathbf{0}.$$

Show that if **a**, **b**, and **c** are linearly independent, then

$$\lambda = -(\mathbf{a} \cdot \mathbf{d})/[\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})], \quad \mu = -(\mathbf{b} \cdot \mathbf{d})/[\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})],$$
$$\nu = -(\mathbf{c} \cdot \mathbf{d})/[\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})].$$

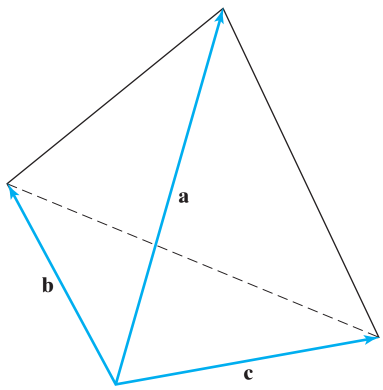


FIGURE 2.22 Tetrahedron.

29. Let **a**, **b**, **c**, and **d** be vectors and λ, μ, ν be scalars satisfying the equation

$$\lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c} + \mathbf{d} = \mathbf{0}.$$

By taking the scalar products of this equation first with $\mathbf{b} \times \mathbf{c}$, then with $\mathbf{a} \times \mathbf{c}$, and finally with $\mathbf{a} \times \mathbf{b}$, show that if **a**, **b**, and **c** are linearly independent, then

$$\lambda = -\mathbf{d} \cdot (\mathbf{b} \times \mathbf{c})/[\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})],$$
$$\mu = -\mathbf{d} \cdot (\mathbf{c} \times \mathbf{a})/[\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})],$$
$$\nu = -\mathbf{d} \cdot (\mathbf{a} \times \mathbf{b})/[\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})].$$

30. Show that $\mathbf{a} = \mathbf{i} + 2\mathbf{j} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} - \mathbf{j} - \mathbf{k}$, and $\mathbf{c} = 4\mathbf{i} + 3\mathbf{j} + \mathbf{ik}$ are linearly independent vectors, and use them with a vector **d** of your choice to verify the results of Exercises 28 and 29.

31. Prove the **Lagrange identity**

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}).$$

2.5 *n*-Vectors and the Vector Space R^n

There are many occasions when it is convenient to generalize a vector and its associated algebra to spaces of more than three dimensions. A typical situation occurs in mechanics, where it is sometimes necessary to consider both the position and the momentum of a particle as functions of time. This leads to the study of a 6-vector, three components of which specify the particle position and three its momentum vector at a time t.

***n*-tuples**

Sets of n numbers $(x_1, x_2, \ldots, x_n)$ in a given order, that can be thought of either as n-vectors or as the coordinates of a point in n-dimensional space are called **ordered *n*-tuples** of real numbers or, simply, ***n*-tuples**.

n-vector

An *n*-vector

If $n \geq 2$ is an integer, and $x_1, x_2, \ldots, x_n$ are real numbers, an ***n*-vector** is an ordered n-tuple

$$(x_1, x_2, \ldots, x_n).$$

components and dimension

The numbers $x_1, x_2, \ldots, x_n$ are called the **components** of the n-vector, x_i is the ith component of the vector, and n is called the **dimension** of the space to which the n-vector belongs. For any given n, the set of all vectors with n real components is called a **real *n*-space** or, simply, an ***n*-space**, and it is denoted by the symbol R^n. A corresponding space exists when the n numbers $x_1, x_2, \ldots, x_n$ are allowed to be complex numbers, leading to a **complex *n*-space** denoted by C^n. In this notation R^3 is the three-dimensional space used in previous sections.

In R^3 the length of a vector was taken as the definition of its norm, so if $\mathbf{r} = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$, then $\|\mathbf{r}\| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$. A generalization of this norm to R^n leads to the following definition.

norm in R^n

The norm in R^n

The **norm** of the n-vector $(x_1, x_2, \ldots, x_n)$, denoted by $\|(x_1, x_2, \ldots, x_n)\|$ is

$$\begin{aligned} \|(x_1, x_2, \ldots, x_n)\| &= \sqrt{(x_1^2 + x_2^2 + \cdots + x_n^2)} \\ &= \left[\sum_{i=1}^{n} x_i^2\right]^{1/2}. \end{aligned} \tag{51}$$

The laws for the equality, addition, and scaling of vectors in R^3 in terms of the components of the vector generalize to R^n as follows.

Equality of *n*-vectors

algebraic rules for equality, addition, and scaling using components

Let $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ be two n-vectors. Then the vectors will be **equal**, written $(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n)$, if, and only if, corresponding components are equal, so that

$$x_1 = y_1, \quad x_2 = y_2, \ldots, x_n = y_n. \tag{52}$$

Addition of *n*-vectors

Let $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ be any two n-vectors. Then the **sum** of these vectors, written $(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n)$, is defined as the vector whose ith component is the sum of the corresponding ith components of the vectors for $i = 1, 2, \ldots, n$, so that

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{53}$$

Scaling an n-vector

Let $(x_1, x_2, \ldots, x_n)$ be an arbitrary n-vector and λ be any scalar. Then the result of **scaling** the vector by λ, written $\lambda(x_1, x_2, \ldots, x_n)$, is defined as the vector whose ith component is λ times the ith component of the original vector, for $i = 1, 2, \ldots, n$, so that

$$\lambda(x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n). \tag{54}$$

The **null (zero)** vector in R^n is the vector $\mathbf{0}$ in which every component is zero, so that

$$\mathbf{0} = (0, 0, \ldots, 0). \tag{55}$$

As with vectors in R^3, so also with n-vectors in R^n, it is convenient to use a single boldface symbol for a vector and the corresponding italic symbols with suffixes when it is necessary to specify the components. So we will write

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad \mathbf{y} = (y_1, y_2, \ldots, y_n).$$

The reasoning that led to the interpretation of Theorem 2.1 on the algebraic rules for the addition and scaling of vectors in R^3 leads also the following theorem for n-vectors.

THEOREM 2.5

Algebraic rules for the addition and scaling of n-vectors in R^n Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be arbitrary n-vectors, and let λ and μ be arbitrary real numbers. Then:

(i) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$;
(ii) $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$;
(iii) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$;
(iv) $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$;
(v) $(\lambda\mu)\mathbf{x} = \lambda(\mu\mathbf{x}) = \mu(\lambda\mathbf{x})$;
(vi) $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$;
(vii) $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$. ∎

Because of this similarity between vectors in R^3 and in R^n, the space R^n is called a **real vector space**, though because the symbol R indicates *real* numbers this is usually abbreviated a **vector space**. Analogously, when the elements of the n-vectors are allowed to be complex, the resulting space is called the **complex vector space** C^n.

So far there would seem to be little difference between vectors in R^3 and R^n, but major differences do exist, and they are best appreciated when geometrical analogies are sought for vector operations in R^n.

dot product of n-vectors

The dot product of n-vectors

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be any two n-vectors. Then the **dot product** of these two vectors, written $\mathbf{x} \cdot \mathbf{y}$ and also called their **inner**

product, is defined as the sum of the products of corresponding components, so that

$$(x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n. \tag{56}$$

The following properties of this dot product are strictly analogous to those of the dot product in R^3 and can be deduced directly from (56).

THEOREM 2.6

Properties of the dot product in R^n Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be any three n-vectors and λ be any scalar. Then:

(i) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$;
(ii) $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$;
(iii) $(\lambda \mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\lambda \mathbf{y}) = \lambda(\mathbf{x} \cdot \mathbf{y})$;
(iv) $\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2$;
(v) $\mathbf{x} \cdot \mathbf{0} = 0$;
(vi) $\|\mathbf{x}\|^2 = 0$ if, and only if, $\mathbf{x} = \mathbf{0}$. ■

The existence of a dot product in R^n allows the Cauchy–Schwarz and triangle inequalities to be generalized, both of which play a fundamental role in the study of vector spaces. Various forms of proof of these inequalities are possible, but the one given here has been chosen because it makes full use of the properties of the dot product listed in Theorem 2.6.

THEOREM 2.7

generalized inequalities for *n*-vectors

The Cauchy–Schwarz and triangle inequalities Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be any two n-vectors. Then

(a) $|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$ (**Cauchy–Schwarz inequality**),

and

(b) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (**triangle inequality**).

Proof We start by proving the Cauchy–Schwarz inequality in (a). The inequality is certainly true if $\mathbf{x} \cdot \mathbf{y} = 0$, so we need only consider the case $\mathbf{x} \cdot \mathbf{y} \neq 0$. Let $\mathbf{x}$ and $\mathbf{y}$ be any two n-vectors, and λ be a scalar. Then, using properties (ii) to (iv) of Theorem 2.6,

$$\begin{aligned} \|\mathbf{x} + \lambda \mathbf{y}\|^2 &= (\mathbf{x} + \lambda \mathbf{y}) \cdot (\mathbf{x} + \lambda \mathbf{y}), \\ &= \|\mathbf{x}\|^2 + \lambda \mathbf{x} \cdot \mathbf{y} + \lambda \mathbf{y} \cdot \mathbf{x} + \lambda^2 \|\mathbf{y}\|^2. \end{aligned}$$

However, by result (1) of Theorem 2.6, $\mathbf{y} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{y}$, so

$$\|\mathbf{x} + \lambda \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\lambda \mathbf{x} \cdot \mathbf{y} + \lambda^2 \|\mathbf{y}\|^2.$$

We now set $\lambda = -\|x\|^2/(\mathbf{x} \cdot \mathbf{y})$ to obtain

$$\|\mathbf{x} + \lambda \mathbf{y}\|^2 = -\|\mathbf{x}\|^2 + (\|\mathbf{x}\|^4 \|\mathbf{y}\|^2)/|\mathbf{x} \cdot \mathbf{y}|^2,$$

where we have used the fact that $(\mathbf{x} \cdot \mathbf{y})^2 = |\mathbf{x} \cdot \mathbf{y}|^2$. As $\|\mathbf{x} + \lambda \mathbf{y}\|^2$ is nonnegative, this result is equivalent to

$$-\|\mathbf{x}\|^2 + (\|\mathbf{x}\|^4 \cdot \|\mathbf{y}\|^2)/|\mathbf{x} \cdot \mathbf{y}|^2 \geq 0.$$

Cancelling the nonnegative number $\|\mathbf{x}\|^2$, which leaves the inequality sign unchanged; rearranging the terms; and taking the square root of the remaining nonnegative result on each side of the inequality yields the Cauchy–Schwarz inequality

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

To prove the triangle inequality (b) we set $\lambda = 1$ and start from the result

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2.$$

As $\mathbf{x} \cdot \mathbf{y}$ may be either positive or negative, $\mathbf{x} \cdot \mathbf{y} \leq |\mathbf{x} \cdot \mathbf{y}|$, so making use of the Cauchy–Schwarz inequality shows that

$$\begin{aligned}\|\mathbf{x} + \mathbf{y}\|^2 &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \cdot \|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.\end{aligned}$$

The *triangle inequality* follows from taking the square root of each side of this inequality, which is permitted because both are nonnegative numbers. ■

The dot product in R^3 allowed the angle between vectors to be determined and, more importantly, it provided a test for the orthogonality of vectors. These same geometrical ideas can be introduced into the vector space R^n if the Cauchy–Schwarz inequality is written in the form

$$-\|\mathbf{x}\| \cdot \|\mathbf{y}\| \leq \mathbf{x} \cdot \mathbf{y} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

After division by the nonnegative number $\|\mathbf{x}\| \cdot \|\mathbf{y}\|$, this becomes

$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \leq 1.$$

This enables the angle θ between the two n-vectors $\mathbf{x}$ and $\mathbf{y}$ to be defined by the result

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}.$$

orthogonality of *n*-vectors

On account of this result, two n-vectors $\mathbf{x}$ and $\mathbf{y}$ in R^n will be said to be **orthogonal** when $\mathbf{x} \cdot \mathbf{y} = 0$.

unit *n*-vector

By analogy with R^3 we will call $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ a **unit *n*-vector** if $\|\mathbf{x}\| = 1$. If we define the unit n-vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ as

$$\mathbf{e}_1 = (1, 0, 0, 0, \ldots, 0),\ \mathbf{e}_2 = (0, 1, 0, 0, \ldots, 0), \ldots, \mathbf{e}_n = (0, 0, 0, 0, \ldots, 1),$$

we see that

$$\mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j, \end{cases}$$

showing that the $\mathbf{e}_i$ are mutually orthogonal unit n-vectors in R^n. As a result of this the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ play the same role in R^n as the vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ in R^3. This allows the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ to be written as

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n,$$

where x_i is the ith component of $\mathbf{x}$.

Now suppose that for $n > 3$, we set

$$\mathbf{u}_1 = (1, 0, 0, 0, \ldots, 0), \quad \mathbf{u}_2 = (0, 1, 0, 0, \ldots, 0), \quad \mathbf{u}_3 = (0, 0, 1, 0, \ldots, 0),$$

and all other $\mathbf{u}_i$ identically zero, so that $\mathbf{u}_i = (0, 0, 0, 0, \ldots, 0)$ for $i = 4, 5, \ldots, n$. Then it is not difficult to see that $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ behave like the unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, so that, in some sense the vector space R^3 is embedded in the vector space R^n with vectors in both spaces obeying the same algebraic rules for addition and scaling. This is recognized by saying that R^3 is a **subspace** of R^n.

subspaces

Subspace of R^n

A subset S of vectors in the vector space R^n is called a **subspace** of R^n if S is itself a vector space that obeys the rules for the addition and the scaling of vectors in R^n.

EXAMPLE 2.19

Find the condition that the set S of vectors of the form $(x, mx + c, 0)$, for any m and all real x forms a subspace of the vector space R^3, and give a geometrical interpretation of the result.

Solution The set S can only contain the null vector $(0, 0, 0)$ if $c = 0$, so if $c \neq 0$ the vectors in S cannot form a subspace of R^3. Now let $c = 0$, so that S contains the null vector. The vector addition law holds, because if $(x, mx, 0)$ and $(x', mx', 0)$ are vectors in S, the sum

$$(x, mx, 0) + (x', mx', 0) = (x + x', m(x + x'), 0)$$

is also a vector in S. The scaling $\lambda(x, mx, 0) = (\lambda x, m\lambda x, 0)$ also generates a vector in S, so the scaling law for vectors also holds, showing that S is a subspace of R^3 provided $c = 0$.

If the three components of vectors in S are regarded as the x-, y-, and z-components of a vector in R^3, the vectors can be interpreted as points on the straight line $y = mx$ passing through the origin and lying in the plane $z = 0$. This subspace is a one-dimensional vector space embedded in the three-dimensional vector space R^3. ■

EXAMPLE 2.20

Test the following subsets of R^n to determine if they form a subspace.

(a) S is the set of vectors $(x_1, x_1 + 1, \ldots, x_n)$ with all the x_i real numbers.

(b) S is the set of vectors $(x_1, x_2, \ldots, x_n)$ with $x_1 + x_2 + \cdots + x_n = 0$ and all the x_i are real numbers.

Solution **(a)** The set S does not contain the null vector and so cannot form a subspace of R^n. This result is sufficient to show that S is not a subspace, but to see what properties of a subspace the set S possesses we consider both the summation and scaling of vectors in S. If $(x_1, x_1 + 1, \ldots, x_n)$ and $(x_1', x_1' + 1, \ldots, x_n')$ are two vectors in S, their sum

$$(x_1, x_1 + 1, \ldots, x_n) + (x_1', x_1' + 1, \ldots, x_n') = (x_1 + x_1', x_1 + x_1' + 2, \ldots, x_n + x_n')$$

is *not* a vector in S, so the summation law is not satisfied.

The scaling condition for vectors is not satisfied, because if λ is an arbitrary scalar,

$$\lambda(x_1, x_1 + 1, \ldots, x_n) = (\lambda x_1, \lambda x_1 + \lambda, \ldots, \lambda x_n) \neq (a, a+1, \ldots), \quad (\lambda n_1 = a)$$

showing that scaling generates another a vector in S. We have proved that the vectors in S do *not* form a subspace of R^n.

(b) The set S does contain the null vector, because $x_1 = x_2 = \cdots = x_n = 0$ satisfies the constraint condition $x_1 + x_2 + \cdots + x_n = 0$. Both the summation law and the scaling law for vectors are easily seen to be satisfied, so this set S does form a subspace of R^n. ■

EXAMPLE 2.21

Let $C(a, b)$ be the space of all real functions of a single real variable x that are continuous for $a < x < b$, and let $S(a, b)$ be the set of all functions belonging to $C(a, b)$ that have a derivative at every point of the interval $a < x < b$. Show that $S(a, b)$ forms a subspace of $C(a, b)$.

Solution In this case a vector in the space is simply any real function of a single real variable x that is continuous in the interval $a < x < b$. The null vector corresponds to the continuous function that is identically zero in the stated interval, so as the derivative of this function is also zero, it follows that the set $S(a, b)$ must also contain the null vector. The sum of continuous functions in $a < x < b$ is a continuous function, and the sum of differentiable functions in this same interval is a differentiable function, so the summation law for vectors is satisfied. Similarly, scaling continuous functions and differentiable functions does not affect either their continuity or their differentiability, so the scaling law for vectors is also satisfied. Thus, $S(a, b)$ forms a subspace of $C(a, b)$. Think of the dimension of these spores as infinite; norm and inner product are easy to define. ■

Summary

This section generalized the concept of a three-dimensional vector to a vector with n components in R^n. It was shown that the magnitude of a vector in three space dimensions generalizes to the norm of a vector in R^n and that in terms of components, the equality, addition, and scaling of vectors in R^n follow the same pattern as with three space dimensions. The dot product was generalized and two fundamental inequalities for vectors in R^n were derived. The concept of orthogonality of vectors was generalized and the notion of a subspace of R^n was introduced.

EXERCISES 2.5

In Exercises 1 through 8 find the sum of the given pairs of vectors, their norms, and their dot product.

1. $(2, 1, 0, 2, 2), (1, -1, 2, 2, 4)$.
2. $(3, -1, -1, 2, -4), (1, 2, 0, 0, 3)$.
3. $(2, 1, -1, 2, 1), (-2, -1, 1, -2, -1)$.
4. $(3, -2, 1, 1, 2, 0, 1), (1, -1, 1, -1, 1, 0, 1)$.
5. $(3, 0, 1, 0), (0, 2, 0, 4)$.
6. $(1, -1, 2, 2, 0, 1), (2, -2, 1, 1, 1, 0)$.
7. $(-1, 2, -4, 0, 1), (2, -1, 1, 0, 2)$.
8. $(3, 1, 2, 4, 1, 1, 1), (1, 2, 3, -1, -2, 1, 3)$.

In Exercises 9 through 12 find the angle between the given pairs of n-vectors and the unit n-vector associated with each vector.

9. $(3, 1, 2, 1), (1, -1, 2, 2)$.
10. $(4, 1, 0, 2), (2, -1, 2, 1)$.
11. $(2, -2, -2, 4), (1, -1, -1, 2)$.
12. $(2, 1, -1, 1), (1, -2, 2, 2)$.

In Exercises 13 through 18 determine if the set of vectors S forms a subspace of the given vector space. Give reasons why S either is or is not a subspace.

13. S is the set of vectors of the form $(x_1, x_2, \ldots, x_n)$ in R^n, with the x_i real numbers and $x_2 = x_1^4$.

14. S is the set of vectors of the form $(x_1, x_2, \ldots, x_n)$ in R^n, with the x_i real numbers and $x_1 + 2x_2 + 3x_3 + \cdots + nx_n = 0$.

15. S is the set of vectors of the form $(x_1, x_2, \ldots, x_n)$ in R^n, with the x_i real numbers and $x_1 + x_2 + x_3 + \cdots + x_n = 2$.

16. S is the set of vectors of the form $(x_1, x_2, \ldots, x_6)$ in R^6, with the x_i real numbers and $x_1 = 0$ or $x_6 = 0$.

17. S is the set of vectors of the form $(x_1, x_2, \ldots, x_6)$ in R^6, with the x_i real numbers and $x_1 - x_2 + x_3 \cdots + x_6 = 0$.

18. S is the set of vectors of the form $(x_1, x_2, \ldots, x_5)$ in R^5, with the x_i real numbers and $x_2 < x_3$.

In Exercises 19 to 23 determine if the given set S is a subspace of the space $C[0, 1]$ of all real valued functions that are continuous on the interval $0 \le x \le 1$. Give reasons why either S is a subspace, or it is not.

19. S is the set of all polynomials of degree two.

20. S is the set of all polynomial functions.

21. S is the set of all continuous functions such that $f(0) = f(1) = 0$.

22. S is the set of all continuous functions such that $f(0) = 0$ and $f(1) = 2$.

23. S is the set of all continuous once differentiable functions such that $f(0) = 0$ and $f'(x) > 0$.

24. Prove that the set S of all vectors lying in any plane in R^3 that passes through the origin forms a subspace of R^3.

25. Explain why the set S of all vectors lying in any plane in R^3 that does not pass through the origin does not form a subspace of R^3.

26. Consider the polynomial $P(\lambda)$ defined as

$$P(\lambda) = \|\mathbf{x} + \lambda\mathbf{y}\|^2,$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors in R^n. Show, provided not both $\mathbf{x}$ and $\mathbf{y}$ are null vectors, that the graph of $P(\lambda)$ as a function of λ is nonnegative, so $P(\lambda) = 0$ cannot have real roots. Use this result to prove the Cauchy–Schwarz inequality

$$|\mathbf{x} \cdot \mathbf{y}| \le \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

27. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in R^n and λ be a scalar. Prove that

$$\|\mathbf{x} + \lambda\mathbf{y}\|^2 + \|\mathbf{x} - \lambda\mathbf{y}\|^2 = 2(\|\mathbf{x}\|^2 + \lambda^2\|\mathbf{y}\|^2).$$

28. If $\mathbf{x}$ and $\mathbf{y}$ are orthogonal vectors in R^n, prove that the Pythagoras theorem takes the form

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2.$$

29. What conditions on the components of vectors $\mathbf{x}$ and $\mathbf{y}$ in the Cauchy–Schwarz inequality cause it to become an equality, so that

$$\sum_{i=1}^{n} x_i y_i = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \left(\sum_{i=1}^{n} y_i^2\right)^{1/2} ?$$

30. Modify the method of proof used in Theorem 2.7 to prove the complex form of the Cauchy–Schwarz inequality

$$\left|\sum_{i=1}^{n} x_i y_i\right| \le \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} + \left(\sum_{i=1}^{n} |y_i|^2\right)^{1/2},$$

where the x_i and y_i are complex numbers.

2.6 Linear Independence, Basis, and Dimension

The concept of the linear independence of a set of vectors in R^3 introduced in Section 2.4 generalizes to R^n and involves a linear combination of n-vectors.

Linear combination of n-vectors

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ be a set of n-vectors in R^n. Then a **linear combination** of the n-vectors is a sum of the form

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_m\mathbf{x}_m,$$

where $c_1, c_2, \ldots, c_m$ are nonzero scalars.

An example of a linear combination of vectors in R^5 is provided by the vector sum ($m = 3, n = 5$)

$$2\mathbf{x}_1 + \mathbf{x}_2 + 3\mathbf{x}_3,$$

where $\mathbf{x}_1 = (1, 2, 3, 0, 4)$, $\mathbf{x}_2 = (2, 1, 4, 1, -3)$, and $\mathbf{x}_3 = (6, 0, 2, 2, -1)$. The vector in R^5 formed by this linear combination is

$$\begin{aligned} 2\mathbf{x}_1 + \mathbf{x}_2 + 3\mathbf{x}_3 &= 2(1, 2, 3, 0, 4) + (2, 1, 4, 1, -3) + 3(6, 0, 2, 2, -1), \\ &= (22, 5, 16, 7, 2). \end{aligned}$$

A linear combination of n-vectors is the most general way of combining n-vectors, and the definition of a linear combination of vectors contains within it the definition of the scaling of a single n-vector as a special case. This can be seen by setting $m = 1$, because this reduces the linear combination to the single scaled n-vector $c_1\mathbf{x}_1$.

linear dependence and independence of *n*-vectors

Linear dependence of *n*-vectors

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ be a set of n-vectors in R^n. Then the set is said to be **linearly dependent** if, and only if, one of the n-vectors can be expressed as a linear combination of the remaining n-vectors.

An example of linear dependence in R^4 is provided by the vectors $\mathbf{x}_1 = (1, 0, 2, 5)$, $\mathbf{x}_2 = (2, 1, 2, 1)$, $\mathbf{x}_3 = (3, 2, 1, 0)$, and $\mathbf{x}_4 = (-1, -1, -1, 7)$, because

$$\mathbf{x}_4 = 2\mathbf{x}_1 - 3\mathbf{x}_2 + \mathbf{x}_3.$$

Linear independence of *n*-vectors

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ be a set of n-vectors in R^n. Then the set is said to be **linearly independent** if, and only if, the n-vectors are not linearly dependent.

A simple example of a set of linearly independent vectors in R^4 is provided by the vectors $\mathbf{e}_1 = (1, 0, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0, 0)$, and $\mathbf{e}_3 = (0, 0, 1, 0)$. The linear independence of these 4-vectors can be seen from the fact that for no choice of c_1 and c_2 can the vector $c_1\mathbf{e}_1 + c_2\mathbf{e}_2$ be made equal to $\mathbf{e}_3$.

To make effective use of the concept of linear independence, and to understand the notion of the *basis* and *dimension* of a vector space, it is necessary to have a test for linear independence. Such a test is provided by the following theorem.

THEOREM 2.8

Linear dependence and independence Let S be a set of non-zero n-vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, with $m \geq 2$. Then:

(a) Set S is linearly dependent if the vector equation

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_m\mathbf{x}_m = \mathbf{0}$$

is true for some set of scalars (constants) $c_1, c_2, \ldots, c_m$ that are not all zero;

(b) Set S is linearly independent if the vector equation

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_m\mathbf{x}_m = \mathbf{0}$$

is only true when $c_1 = c_2 = \cdots = c_m = 0$.

Proof To establish result (a) it is necessary to show that the conditions of the definition of linear dependence are satisfied. First, if the set S of n-vectors is linearly dependent, scalars $d_1, d_2, \ldots, d_m$ exist such that

$$d_1\mathbf{x}_1 + d_2\mathbf{x}_2 + \cdots + d_m\mathbf{x}_m = \mathbf{0}.$$

There is no loss of generality in assuming that $d_1 \neq 0$, because if this is not the case a renumbering of the vectors can always make this possible. Consequently,

$$\mathbf{x}_1 = (-d_2/d_1)\mathbf{x}_2 + (-d_3/d_1)\mathbf{x}_3 + \cdots + (-d_m/d_1)\mathbf{x}_m,$$

which shows, as claimed, that the set S is linearly dependent, because $\mathbf{x}_1$ is linearly dependent on $\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_m$. A similar argument applies to show that $\mathbf{x}_r$ is linearly dependent on the remaining n-vectors in S provided $d_r \neq 0$, for $r = 2, 3, \ldots, m$.

Conversely, if one of the n-vectors in set S, say $\mathbf{x}_1$, is linearly dependent on the remaining n-vectors in the set, scalars $d_1, d_2, \ldots, d_m$ can be found such that

$$\mathbf{x}_1 = d_2\mathbf{x}_2 + \cdots + d_m\mathbf{x}_m,$$

so that

$$\mathbf{x}_1 - d_2\mathbf{x}_2 - \cdots - d_m\mathbf{x}_m = 0.$$

This result is of the form given in definition of linear dependence with $c_1 = 1$, $c_2 = -d_2, \ldots, c_m = -d_m$, not all of which constants are zero, so again the set of n-vectors in S is seen to be linearly dependent.

To establish result (b), suppose, if possible, that the set S of vectors is linearly independent, but that some scalars $d_1, d_2, \ldots, d_m$ that are not all zero can be found such that

$$d_1\mathbf{x}_1 + d_2\mathbf{x}_2 + \cdots + d_m\mathbf{x}_m = \mathbf{0}.$$

Then if $d_1 \neq 0$, say, is one of these scalars, it follows that

$$\mathbf{x}_1 = (-d_2/d_1)\mathbf{x}_2 + (-d_3/d_1)\mathbf{x}_3 + \cdots + (-d_m/d_1)\mathbf{x}_m,$$

which is impossible because this shows that, contrary to the hypothesis, $\mathbf{x}_1$ is linearly dependent on the remaining n-vectors in S. So we must have $c_1 = c_2 = \cdots = c_m = 0$. ■

A systematic and efficient computational method for the application of Theorem 2.8 to vectors in R^n will be developed in the next chapter for the three separate cases that arise, (a) $m < n$, (b) $m = n$, and (c) $m > n$. However, when n and m are small, a straightforward approach is possible, as illustrated in the next example.

EXAMPLE 2.22

Test the following sets of vectors in R^4 for linear dependence or independence.

(a) $\mathbf{x}_1 = (2, 1, 1, 0), \quad \mathbf{x}_2 = (0, 2, 0, 1), \quad \mathbf{x}_3 = (1, 1, 0, 2), \quad \mathbf{x}_4 = (0, 2, 1, 1).$

(b) $\mathbf{x}_1 = (4, 0, 2), \quad \mathbf{x}_2 = (2, 2, 0), \quad \mathbf{x}_3 = (1, 1, 0), \quad \mathbf{x}_4 = (5, 1, 2).$

Solution In both (a) and (b) it is necessary to consider the vector equation

$$c_1\mathbf{x}_1 + \cdots + c_m\mathbf{x}_m = \mathbf{0}.$$

If the equation is only satisfied when $c_1 = c_2 = \cdots = c_m = 0$, the set of vectors will be linearly independent, whereas if a solution can be found in which not all of the constants c_1, c_2, c_3, c_4 vanish, the set of vectors will be linearly dependent.

(a) Substituting for $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ in the preceding equation and equating corresponding components show the coefficients c_i must satisfy the following equations

$$\begin{aligned} 2c_1 + c_3 &= 0 \\ c_1 + 2c_2 + c_3 + 2c_4 &= 0 \\ c_1 + c_4 &= 0 \\ c_2 + 2c_3 + c_4 &= 0. \end{aligned}$$

The third equation shows that $c_4 = -c_1$, so the equations can be rewritten as

$$\begin{aligned} 2c_1 + c_3 &= 0 \\ -c_1 + 2c_2 + c_3 &= 0 \\ c_2 - c_1 + 2c_3 &= 0. \end{aligned}$$

Adding twice the third equation to the first equation shows that $c_3 = 0$, so $c_1 = 0$, and it then follows that $c_2 = c_3 = c_4 = 0$. This has established the linear independence of the set of vectors in (a).

(b) Proceeding in the same manner with the set of vectors in (b) leads to the following equations for the coefficients c_i:

$$\begin{aligned} 4c_1 + 2c_2 + c_3 + 5c_4 &= 0 \\ 2c_2 + c_3 + c_4 &= 0 \\ 2c_1 + 2c_4 &= 0. \end{aligned}$$

The third equation shows that $c_4 = -c_1$, so using this result in the first two equations reduces the first one to

$$-c_1 + 2c_2 + c_3 = 0$$

and the second to

$$-c_1 + 2c_2 + c_3 = 0.$$

There is only one equation connecting c_1, c_2, and c_3, and hence also c_4. This means that if c_2 and c_3 are given arbitrary values, not both of which are zero, the constants c_1 and c_4 will be determined in terms of them. Thus, a set of constants c_1, c_2, c_3, c_4 that are not all zero can be found that satisfy the vector equation, showing that the set of vectors in (b) is linearly dependent. This set of constants is not unique, but this does affect the conclusion that the set of vectors is linearly dependent, because to establish linear dependence it is sufficient that at least one such set of constants can be found. ■

Example 2.22 has shown one way in which Theorem 2.8 can be implemented for vectors in R^n, but it also illustrates the need for a systematic approach to the solution of the system of equations for the coefficients when n is large.

A trivial case of Theorem 2.8 arises when the set of vectors S contains the null vector $\mathbf{0}$, because then the set of vectors in S is always linearly dependent. This can be seen by assuming that $\mathbf{x}_1 = \mathbf{0}$, because then the vector equation in the theorem becomes

$$c_1\mathbf{0} + c_2\mathbf{x}_2 + \cdots + c_m\mathbf{x}_m = \mathbf{0}.$$

This vector equation is satisfied if $c_1 \neq 0$ (arbitrary) and $c_2 = c_3 = \cdots = c_m = 0$, so, as not all of the coefficients are zero, the set of vectors must be linearly dependent.

We conclude this introduction to the vector space R^n by defining the *span*, a *basis*, and the *dimension* of a vector space.

span of a vector space

Span of a vector space

Let the set of non-zero vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ belonging to a vector space V have the property that every vector in V can be expressed as a linear combination of these vectors. Then the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ are said to **span** the vector space V.

EXAMPLE 2.23

All vectors $\mathbf{v}$ in the (x, y)-plane are spanned by the vectors $\mathbf{i}$ and $\mathbf{j}$, because any vector $\mathbf{v} = (v_1, v_2)$ can always be written $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j}$. This is an example of vectors spanning the space R^2. ■

EXAMPLE 2.24

The vector space R^n is spanned by the unit n-vectors

$$\mathbf{e}_1 = (1, 0, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, 0, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, 0, 0, \ldots, 1).$$ ■

EXAMPLE 2.25

The subspace R^3 of the vector space R^5 is spanned by the unit vectors

$$\mathbf{e}_1 = (1, 0, 0, 0, 0), \quad \mathbf{e}_2 = (0, 1, 0, 0, 0), \quad \mathbf{e}_3 = (0, 0, 1, 0, 0),$$

because all vectors $\mathbf{v} = (v_1, v_2, v_3)$ in R^3 can be written in the form of the linear combination $\mathbf{v} = v_1\mathbf{e}_1 + v_1\mathbf{e}_2 + v_3\mathbf{e}_3$. ■

basis of a vector space in R^n

Basis of a vector space

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be vectors in R^n. Then the vectors are said to form a **basis** for the vector space R^n if:

(i) The vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent.
(ii) Every vector in R^n can be expressed as a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

dimension of a vector space

Dimension of a vector space

The **dimension** of a vector space is the number of vectors in its basis.

EXAMPLE 2.26 A basis for the space of ordinary vectors in three dimensions is provided by the vectors **i**, **j**, and **k**, so the dimension of the space is 3. ■

EXAMPLE 2.27 A basis for R^n is provided by the n vectors

$$\mathbf{e}_1 = (1, 0, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, 0, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, 0, 0, \ldots, 1),$$

so its dimension is n. ■

EXAMPLE 2.28 It was shown in Example 2.20 (b) that the set S of vectors $(x_1, x_2, \ldots, x_n)$ with $x_1 + x_2 + \cdots + x_n = 0$ forms a subspace of R^n. The dimension of R^n is n, but the constraint condition $x_1 + x_2 + \cdots + x_n = 0$ implies that only $n - 1$ of the components $x_1, x_2, \ldots, x_n$ can be specified independently, because the constraint itself determines the value of the remaining component. This in turn implies that the basis for the subspace S can only contain $n - 1$ linearly independent vectors, so S must have dimension $n - 1$. ■

More information on linear vector spaces can be found in references [2.1] and [2.5] to [2.12].

Summary

In this section the concepts of linear dependence and independence were generalized to vectors in R^n, and the span of a vector space was defined as a set of vectors in R^n with the property that every vector in R^n can be expressed as a linear combination of these vectors. Naturally in R^n, as in R^3, a set of vectors spanning the space is not unique. The smallest set of n vectors spanning a vector space is said to form a basis for the vector space, and the dimension of a vector space is the number of vectors in its basis. This corresponds to the fact that the unit vectors **i**, **j**, and **k** form a basis for the ordinary three-dimensional space R^3, because every vector in this space can be represented as a linear combination of **i**, **j**, and **k**.

EXERCISES 2.6

In Exercises 1 through 12 determine if the set of m vectors in three-dimensional space is linearly independent by solving for the scalars $c_1, c_2 \ldots c_m$ in Theorem 2.8. Where appropriate, verify the result by using Theorem 2.3.

1. $\mathbf{a} = \mathbf{i} + 2\mathbf{j} + \mathbf{k},\ \mathbf{b} = \mathbf{i} - \mathbf{j} + \mathbf{k},\ \mathbf{c} = 2\mathbf{i} + \mathbf{k}$.
2. $\mathbf{a} = 3\mathbf{i} - \mathbf{j} + \mathbf{k},\ \mathbf{b} = \mathbf{i} + 3\mathbf{k},\ \mathbf{c} = 5\mathbf{i} - \mathbf{j} + 7\mathbf{k}$.
3. $\mathbf{a} = 2\mathbf{i} - \mathbf{j} + \mathbf{k},\ \mathbf{b} = 3\mathbf{i} + \mathbf{j} - \mathbf{k},\ \mathbf{c} = 8\mathbf{i} + \mathbf{j} + 7\mathbf{k}$.
4. $\mathbf{a} = 3\mathbf{i} + 2\mathbf{k},\ \mathbf{b} = \mathbf{i} + \mathbf{j} + 2\mathbf{k},\ \mathbf{c} = 11\mathbf{i} + 2\mathbf{j} - 2\mathbf{k}$.
5. $\mathbf{a} = 4\mathbf{i} - \mathbf{j} + 3\mathbf{k},\ \mathbf{b} = \mathbf{i} + 4\mathbf{j} - 2\mathbf{k},\ \mathbf{c} = 3\mathbf{i} - \mathbf{j} - \mathbf{k}$.
6. $\mathbf{a} = \mathbf{i} + \mathbf{j} - \mathbf{k},\ \mathbf{b} = \mathbf{i} - \mathbf{j} + \mathbf{k},\ \mathbf{c} = -\mathbf{i} + \mathbf{j} + \mathbf{k}$.
7. $\mathbf{a} = \mathbf{i} + 2\mathbf{j} + \mathbf{k},\ \mathbf{b} = \mathbf{i} + 3\mathbf{j} - \mathbf{k},\ \mathbf{c} = 3\mathbf{i} + 10\mathbf{j} - 5\mathbf{k}$.
8. $\mathbf{a} = 2\mathbf{i} + 3\mathbf{j} + \mathbf{k},\ \mathbf{b} = \mathbf{i} - 3\mathbf{j} + 2\mathbf{k},\ \mathbf{c} = \mathbf{i} + 15\mathbf{j} - 4\mathbf{k}$.
9. $\mathbf{a} = 3\mathbf{i} - \mathbf{j} + 2\mathbf{k},\ \mathbf{b} = \mathbf{i} + \mathbf{j} + \mathbf{k}\quad (m = 2)$.
10. $\mathbf{a} = \mathbf{i} + \mathbf{j} + \mathbf{k},\ \mathbf{b} = \mathbf{i} + 2\mathbf{j} + \mathbf{k},\ \mathbf{c} = \mathbf{i} + 3\mathbf{j} + \mathbf{k},\ \mathbf{d} = \mathbf{i} - 4\mathbf{j} + \mathbf{k}$ $(m = 4)$.
11. $\mathbf{a} = \mathbf{i} - \mathbf{j} + 3\mathbf{k},\ \mathbf{b} = 2\mathbf{i} - \mathbf{j} + 2\mathbf{k},\ \mathbf{c} = \mathbf{i} + \mathbf{k},\ \mathbf{d} = 3\mathbf{i} + \mathbf{j} + \mathbf{k}$ $(m = 4)$.
12. $\mathbf{a} = \mathbf{i} + \mathbf{j},\ \mathbf{b} = \mathbf{j} + \mathbf{k},\ \mathbf{c} = \mathbf{i} - \mathbf{k}$.

In Exercises 13 through 16, determine if the set of vectors in R^4 is linearly independent by using the method of Example 2.22.

13. $(1, 3, -1, 0), (1, 2, 0, 1), (0, 1, 0, -1), (1, 1, 0, 1)$.
14. $(1, -2, 1, 2), (4, -1, 0, 2), (2, 1, -1, 1), (1, 0, 0, -1)$.
15. $(2, 1, 0, 1), (1, 0, 1, 1), (4, 1, 2, -1), (1, 0, 1, -1)$.
16. $(1, 2, 1, 1), (1, -2, 0, -1), (1, 1, 1, 2), (1, -1, 0, 0)$.

In Exercises 17 through 20, find a basis and the dimension of the given subspace S.

17. The subspace S of vectors in R^5 of the form $(x_1, x_2, x_3, x_4, x_5)$ with $x_1 = x_2$.
18. The subspace S of vectors in R^4 of the form (x_1, x_2, x_3, x_4) with $x_1 = 2x_2$.
19. The subspace S of vectors in R^5 of the form $(x_1, x_2, x_3, x_4, x_5)$ with $x_1 = x_2 = 2x_3$.

20. The subspace S of vectors in R^6 of the form $(x_1, x_2, x_3, x_4, x_5, x_6)$ with $x_1 = 2x_2$ and $x_3 = -x_4$.

21. Let $\mathbf{u} = \cos^2 x$ and $\mathbf{v} = \sin^2 x$ form a basis for a vector space V. Find which of the following can be represented in terms of $\mathbf{u}$ and $\mathbf{v}$, and so lie in V.

(a) 2. (b) $\sin 2x$. (c) 0. (d) $\cos 2x$. (e) $2 + 3x$. (f) $3 - 4\cos 2x$.

22. Given that $r \le n$, prove that any subset S of r vectors selected from a set of n linearly independent vectors is linearly independent.

2.7 Gram–Schmidt Orthogonalization Process

A set of vectors forming a basis for a vector space is not unique, and having obtained a basis by some means, it is often useful to replace it by an equivalent set of orthogonal vectors. The **Gram–Schmidt orthogonalization process** accomplishes this by means of a sequence of simple steps that have a convenient geometrical interpretation. We now develop the Gram–Schmidt orthogonalization process for geometrical vectors in R^3, though in Section 4.2 the method will be extended to vectors in R^n to enable orthogonal matrices to be constructed from a set of eigenvectors associated with a symmetric matrix.

Let us now show how any basis for R^3, comprising three nonorthogonal linearly independent vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, can be used to construct an equivalent basis involving three linearly independent orthogonal vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$. It is essential that the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ be linearly independent, because if not, the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ generated by the Gram–Schmidt orthogonalization process will be linearly dependent and so cannot form a basis for R^3. The derivation of the method starts by setting

$$\mathbf{u}_1 = \mathbf{a}_1,$$

where the choice of $\mathbf{a}_1$ instead of $\mathbf{a}_2$ or $\mathbf{a}_3$ is arbitrary.

The component of $\mathbf{a}_2$ in the direction of $\mathbf{u}_1$ is $\mathbf{u}_1 \cdot \mathbf{a}_2$, so the vector component of $\mathbf{a}_2$ in this direction is

$$(\mathbf{u}_1 \cdot \mathbf{a}_2)\mathbf{u}_1 = \frac{(\mathbf{u}_1 \cdot \mathbf{a}_2)\mathbf{u}_1}{\|\mathbf{u}_1\|^2},$$

and this always exists because $\|\mathbf{u}_1\|^2 > 0$. Subtracting this vector from $\mathbf{a}_2$ gives a vector $\mathbf{u}_2$ that is normal to $\mathbf{u}_1$, so

$$\mathbf{u}_2 = \mathbf{a}_2 - \frac{(\mathbf{u}_1 \cdot \mathbf{a}_2)\mathbf{u}_1}{\|\mathbf{u}_1\|^2}.$$

Similarly, to find a vector normal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ involving $\mathbf{a}_3$, it is necessary to subtract from $\mathbf{a}_3$ the components of vector $\mathbf{a}_3$ in the direction of $\mathbf{u}_1$ and also in the direction of $\mathbf{u}_2$, so that

$$\mathbf{u}_3 = \mathbf{a}_3 - \frac{(\mathbf{u}_1 \cdot \mathbf{a}_3)\mathbf{u}_1}{\|\mathbf{u}_1\|^2} - \frac{(\mathbf{u}_2 \cdot \mathbf{a}_3)\mathbf{u}_2}{\|\mathbf{u}_2\|^2},$$

and this also always exists, because $\|\mathbf{u}_1\|^2 > 0$ and $\|\mathbf{u}_2\|^2 > 0$.

If an orthonormal basis is required, it is necessary to normalize the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ by dividing each by its norm.

Rule for the Gram–Schmidt orthogonalization process in R^3

A set of nonorthogonal linearly independent vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ that form a basis in R^3 can be used to generate an equivalent orthogonal basis involving the vectors, $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ by setting

$$\mathbf{u}_1 = \mathbf{a}_1, \quad \mathbf{u}_2 = \mathbf{a}_2 - \frac{(\mathbf{u}_1 \cdot \mathbf{a}_2)\mathbf{u}_1}{\|\mathbf{u}_1\|^2}, \quad \text{and}$$

$$\mathbf{u}_3 = \mathbf{a}_3 - \frac{(\mathbf{u}_1 \cdot \mathbf{a}_3)\mathbf{u}_1}{\|\mathbf{u}_1\|^2} - \frac{(\mathbf{u}_2 \cdot \mathbf{a}_3)\mathbf{u}_2}{\|\mathbf{u}_2\|^2}.$$

As already remarked, the choice of $\mathbf{a}_1$ as the vector with which to start the orthogonalization process was arbitrary, and the process could equally well have been started by setting $\mathbf{u}_1 = \mathbf{a}_2$ or $\mathbf{u}_1 = \mathbf{a}_3$. Using a different vector will produce a different set of orthogonal vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$, but any basis for R^3 is equivalent to any other basis, so unless there is a practical reason for starting with a particular vector, the choice is immaterial.

EXAMPLE 2.29 Given the nonorthogonal basis $\mathbf{a}_1 = \mathbf{i} - \mathbf{j} - \mathbf{k}$, $\mathbf{a}_2 = \mathbf{i} + \mathbf{j} + \mathbf{k}$, and $\mathbf{a}_3 = -\mathbf{i} + 2\mathbf{k}$, use the Gram–Schmidt orthogonalization process to find an equivalent orthogonal basis, and then find the corresponding orthonormal basis.

Solution Using the preceding rule we start with $\mathbf{u}_1 = \mathbf{i} - \mathbf{j} - \mathbf{k}$, and to find $\mathbf{u}_2$ we need to use the results $\mathbf{u}_1 \cdot \mathbf{a}_2 = -1$ and $\|\mathbf{u}_1\|^2 = 3$, so that

$$\mathbf{u}_2 = \mathbf{i} + \mathbf{j} + \mathbf{k} - (-1/3)(\mathbf{i} - \mathbf{j} - \mathbf{k}) = (4/3)\mathbf{i} + (2/3)\mathbf{j} + (2/3)\mathbf{k}.$$

To find $\mathbf{u}_3$ we need to use the results $\mathbf{u}_1 \cdot \mathbf{a}_3 = -3$, $\|\mathbf{u}_1\|^2 = 3$, $\mathbf{u}_2 \cdot \mathbf{a}_3 = 0$, and $\|\mathbf{u}_2\|^2 = 24/9$, so that

$$\mathbf{u}_3 = -\mathbf{i} + 2\mathbf{k} - (-3/3)(\mathbf{i} - \mathbf{j} - \mathbf{k}) = -\mathbf{j} + \mathbf{k}.$$

So the required equivalent orthogonal basis is

$$\mathbf{u}_1 = \mathbf{i} - \mathbf{j} - \mathbf{k}, \quad \mathbf{u}_2 = (4/3)\mathbf{i} + (2/3)\mathbf{j} + 2/3\mathbf{k}, \quad \text{and} \quad \mathbf{u}_3 = -\mathbf{j} + \mathbf{k}.$$

The corresponding orthonormal basis obtained by dividing each of these vectors by its norm (modulus) is

$$\hat{\mathbf{u}}_1 = (1/\sqrt{3})\mathbf{u}_1, \quad \hat{\mathbf{u}}_2 = (1/2)\sqrt{(3/2)}\mathbf{u}_2 \quad \text{and} \quad \hat{\mathbf{u}}_3 = (1/\sqrt{2})\mathbf{u}_3. \qquad \blacksquare$$

Other accounts of the Gram–Schmidt orthogonalization process are to be found in references [2.1] and [2.7] to [2.12].

Summary

In this section it is shown how in R^3 the Gram–Schmidt orthogonalization process converts any three nonorthogonal linearly independent vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ into three orthogonal vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$. If necessary, the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ can then be normalized in the usual manner to form an orthogonal set of unit vectors.

EXERCISES 2.7

In Exercises 1 through 6, use the given nonorthogonal basis for vectors in R^3 to find an equivalent orthogonal basis by means of the Gram–Schmidt orthogonalization process.

1. $\mathbf{a}_1 = \mathbf{i} + 2\mathbf{j} + \mathbf{k}$, $\mathbf{a}_2 = \mathbf{i} - \mathbf{j}$, $\mathbf{a}_3 = 2\mathbf{j} - \mathbf{k}$.

2. $\mathbf{a}_1 = \mathbf{j} + 3\mathbf{k}$, $\mathbf{a}_2 = \mathbf{i} + \mathbf{j} - \mathbf{k}$, $\mathbf{a}_3 = \mathbf{i} + 2\mathbf{k}$.

3. $\mathbf{a}_1 = 2\mathbf{i} + \mathbf{j}$, $\mathbf{a}_2 = 2\mathbf{j} + \mathbf{k}$, $\mathbf{a}_3 = \mathbf{k}$.

4. $\mathbf{a}_1 = \mathbf{i} + 3\mathbf{k}$, $\mathbf{a}_2 = \mathbf{i} - \mathbf{j} + \mathbf{k}$, $\mathbf{a}_3 = 2\mathbf{i} + \mathbf{j}$.

5. $\mathbf{a}_1 = -\mathbf{i} + \mathbf{k}$, $\mathbf{a}_2 = 2\mathbf{j} + \mathbf{k}$, $\mathbf{a}_3 = \mathbf{i} + \mathbf{j} + \mathbf{k}$.

6. $\mathbf{a}_1 = \mathbf{i} + \mathbf{k}$, $\mathbf{a}_2 = -\mathbf{j} + \mathbf{k}$, $\mathbf{a}_3 = \mathbf{i} + \mathbf{j} + 2\mathbf{k}$.

In Exercises 7 and 8, find two different but equivalent sets of orthogonal vectors by arranging the same three nonorthogonal vectors in the orders indicated.

7. **(a)** $\mathbf{a}_1 = 3\mathbf{j} - \mathbf{k}$, $\mathbf{a}_2 = \mathbf{i} + \mathbf{j}$, $\mathbf{a}_3 = \mathbf{i} + 2\mathbf{k}$.

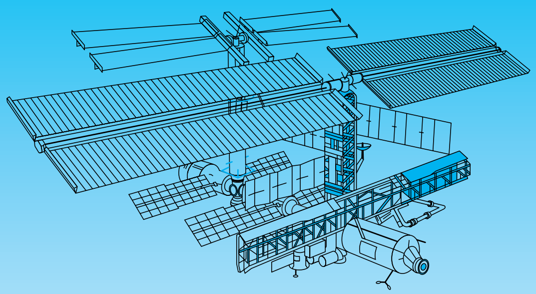
(b) $\mathbf{a}_1 = \mathbf{i} + \mathbf{j}$, $\mathbf{a}_2 = 3\mathbf{j} - \mathbf{k}$, $\mathbf{a}_3 = \mathbf{i} + 2\mathbf{k}$.

8. **(a)** $\mathbf{a}_1 = \mathbf{j} - \mathbf{k}$, $\mathbf{a}_2 = \mathbf{i} + \mathbf{k}$, $\mathbf{a}_3 = -\mathbf{i} - \mathbf{j} + \mathbf{k}$.

(b) $\mathbf{a}_1 = -\mathbf{i} - \mathbf{j} + \mathbf{k}$, $\mathbf{a}_2 = \mathbf{i} + \mathbf{k}$, $\mathbf{a}_3 = \mathbf{j} - \mathbf{k}$.

This Page Intentionally Left Blank

CHAPTER 3



Matrices and Systems of Linear Equations

Many types of problems that arise in engineering and physics give rise to linear algebraic simultaneous equations. A typical engineering example involves the determination of the forces acting in the struts of a pin-jointed structure like a truss that forms the side of a bridge supporting a load. The determination of the forces in a strut is important in order to know when it is in compression or tension, and to ensure that no truss exceeds its safe load. The analysis of the forces in structures of this type gives rise to a set of linear simultaneous equations that relate the forces in the struts and the external load.

It is necessary to know when systems of linear equations are consistent so a solution exists, when they are inconsistent so there is no solution, and whether when a solution exists it is unique or nonunique in the sense that it involves a number of arbitrary parameters. In practical problems all of these mathematical possibilities have physical meaning, and in the case of a truss, the inability to determine the forces acting in a particular strut indicates that it is redundant and so can be removed without compromising the integrity of the structure.

A more complicated though very similar situation occurs when linearly vibrating systems are coupled together, as may happen when an active vibration damper is attached to a spring-mounted motor. However, in this case it is a system of simultaneous linear ordinary differential equations determining the amplitudes of the vibrations of the motor and vibration damper that are coupled together. The analysis of this problem, which will be considered later, also gives rise to a linear system of simultaneous algebraic equations.

Linear ordinary differential equations are also coupled together when working with linear control systems involving feedback. When such systems are solved by means of the Laplace transform to be described later, linear algebraic systems again arise and the nature of the zeros of the determinant of a certain quantity then determines the stability of the control system.

Linear systems of simultaneous algebraic equations also play an essential role in computer graphics, where at the simplest level they are used to transform images by translating, rotating, and stretching them by differing amounts in different directions.

Although each equation in a system of linear algebraic equations can be considered separately, such can be discovered about the properties of the physical problem that gave rise to the equations if the system of equations can be studied as a whole. This can be accomplished by using the algebra of matrices that provides a way of analyzing systems

as a single entity, and it is the purpose of this chapter to introduce and develop this aspect of what is called linear algebra.

After defining the notion of a matrix, this chapter develops the fundamental matrix operations of equality, addition, scaling, transposition, and multiplication. Various applications of matrices are given, and the brief review of determinants given in Chapter 1 is developed in greater detail, prior to its use when considering the solution of systems of linear algebraic equations.

The concept of elementary row operations is introduced and used to reduce systems of linear algebraic equations to a form that shows whether or not a unique solution exists. When a solution does exist, which is either unique or determined in terms of some of the remaining variables, this reduction enables the solution to be found immediately.

The inverse of an $n \times n$ matrix is defined and shown only to exist when the determinant of the matrix is nonvanishing, and, finally, the derivative of a matrix whose elements are functions of a variable is introduced and some of its most important properties are derived.

3.1 Matrices

Matrices arise naturally in many different ways, one of the most common being in the study of systems of linear equations such as

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots& \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{1}$$

In system (1) the numbers a_{ij} are the **coefficients** of the equations, the numbers b_i are the **nonhomogeneous terms**, and the number of equations m may equal, exceed, or be less than n, the number of unknowns $x_1, x_2, \ldots, x_n$.

System (1) is said to be **homogeneous** when $b_1 = b_2 = \cdots = b_m = 0$, and to be **nonhomogeneous** when at least one of the b_i is nonvanishing. The algebraic properties of the system are determined by the array of coefficients a_{ij}, the nonhomogeneous terms b_i and the numbers m and n. From now on, the array of coefficients and the nonhomogeneous terms on the right will be denoted by the single symbols $\mathbf{A}$ and $\mathbf{b}$, respectively, where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \tag{2}$$

The array of mn coefficients a_{ij} in m rows and n columns that form $\mathbf{A}$ is an example of an $m \times n$ **matrix**, where $m \times n$ is read "m by n." The array $\mathbf{b}$ is an example of an $m \times 1$ **matrix**, and it is called an m element **column vector**. We will use the convention that an array such as $\mathbf{A}$, with two or more rows and two or more columns, will be denoted by a boldface capital letter. An array with a single row, or a column such as $\mathbf{b}$, will be denoted by a boldface lowercase letter.

Each entry in a matrix is called an **element** of the matrix, and entries may be numbers, functions, or even matrices themselves. The suffixes associated with an element show its position in the matrix, because the first suffix is the **row number**

and the second is the **column number**. Because of this convention, the element a_{35} in a matrix belongs to the third row and the fifth column of the matrix. So, for example, if $\mathbf{A}$ is a 3×2 matrix and its general element $a_{ij} = i + 3j$, then as i may only take the values 1, 2, and 3, and j the values 1 and 2, it follows that

$$\mathbf{A} = \begin{bmatrix} 4 & 7 \\ 5 & 8 \\ 6 & 9 \end{bmatrix}.$$

In a *column* vector $\mathbf{c}$ with elements $c_{11}, c_{21}, c_{31}, \ldots, c_{m1}$, as only a single column is involved, it is usual to vary the suffix convention by omitting the second suffix and instead numbering the elements sequentially as $c_1, c_2, c_3, \ldots, c_m$, so that

$$\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_m \end{bmatrix}.$$

Later it will be necessary to introduce **row vectors**, and in an s element row vector $\mathbf{r}$ with elements $r_{11}, r_{12}, r_{13}, \ldots, r_{1s}$, the notation is again simplified, this time by omitting the first suffix and numbering the elements sequentially as $r_1, r_2, \ldots, r_s$, so

$$\mathbf{r} = [r_1,\ r_2, \ldots, r_s]. \tag{3}$$

In general, row and column vectors will be denoted by boldface lowercase letters such as $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{x}$, and matrices such as the coefficient matrix in (2) will be denoted by boldface capital letters such as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{P}$, and $\mathbf{Q}$.

A different convention that is also used to denote a matrix involves enclosing the array between curved brackets instead of the square ones used here. Thus,

$$\begin{pmatrix} 1 & 5 & 9 \\ -3 & 2 & 4 \end{pmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 5 & 9 \\ -3 & 2 & 4 \end{bmatrix} \tag{4}$$

denote the same 2×3 matrix. A matrix should never be enclosed between two vertical rules in order to avoid confusion with the determinant notation because

$$\begin{bmatrix} 3 & -4 \\ 5 & 2 \end{bmatrix} \quad \text{is a matrix, but} \quad \begin{vmatrix} 3 & -4 \\ 5 & 2 \end{vmatrix} = 26 \quad \text{is a determinant.}$$

Definition of a matrix

An $m \times n$ **matrix** is an array of mn entries, called **elements**, arranged in m rows and n columns. If a matrix is denoted by $\mathbf{A}$, then the element in its ith row and jth column is denoted by a_{ij} and

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}.$$

some typical matrices

The following are typical examples of matrices:

A 1×1 matrix: $[3]$; a single element may be regarded as a matrix.

A 3×4 matrix: $\begin{bmatrix} 1 & 3 & 5 & 0 \\ 2 & -1 & 4 & 3 \\ 7 & 2 & 1 & 6 \end{bmatrix}$; a matrix with real numbers as elements.

A 2×2 matrix: $\begin{bmatrix} 1+i & 1-i \\ 3+4i & 2-3i \end{bmatrix}$; a matrix with complex numbers as elements.

A 2×2 matrix: $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$; a matrix with functions as elements.

A 1×3 matrix: $[2, \ -5, \ 7]$; a three-element row vector.

A 2×1 matrix: $\begin{bmatrix} 11 \\ 9 \end{bmatrix}$; a two-element column vector.

A **square matrix** is a matrix in which the number of rows m equals the number of columns n. A typical square matrix is the 3×3 matrix

$$\begin{bmatrix} 2 & 0 & 5 \\ 1 & -3 & 4 \\ 3 & 1 & 7 \end{bmatrix}.$$

Definition of the equality of matrices

Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix and $\mathbf{B} = [b_{ij}]$ be a $p \times q$ matrix. Then matrices $\mathbf{A}$ and $\mathbf{B}$ will be **equal**, written $\mathbf{A} = \mathbf{B}$, if, and only if:

(a) $\mathbf{A}$ and $\mathbf{B}$ have the same number of rows, and the same number of columns, so that $m = p$ and $n = q$, and

(b) $a_{ij} = b_{ij}$, for each i and j.

Equality of matrices means that if $\mathbf{A}$ and $\mathbf{B}$ are equal, then each is an identical copy of the other.

EXAMPLE 3.1

If $\mathbf{A} = \begin{bmatrix} 2 & 3 & a \\ b & 6 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & 3 & 9 \\ -3 & 6 & 1 \end{bmatrix}$, and $\mathbf{C} = \begin{bmatrix} 2 & 3 & 9 \\ -3 & 6 & 1 \\ 0 & 0 & 0 \end{bmatrix}$,

then $\mathbf{A} = \mathbf{B}$ if and only if $a = 9$ and $b = -3$, but $\mathbf{A} \neq \mathbf{C}$ and $\mathbf{B} \neq \mathbf{C}$. ■

Definition of matrix addition

The addition of matrices $\mathbf{A}$ and $\mathbf{B}$ is only defined if the matrices each have the same number of rows and the same number of columns. Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ be $m \times n$ matrices. Then the the $m \times n$ matrix formed by adding $\mathbf{A}$ and $\mathbf{B}$, called the **sum** of $\mathbf{A}$ and $\mathbf{B}$ and written $\mathbf{A} + \mathbf{B}$, is the matrix whose element in the ith row and jth column is $a_{ij} + b_{ij}$, for each i and j, so that

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}].$$

Matrices that can be added are said to be **conformable** for addition.

It is an immediate consequence of this definition that $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, so matrix addition is **commutative**.

Definition of the transpose of a matrix

Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix. Then the **transpose** of $\mathbf{A}$, denoted by $\mathbf{A}^{\mathrm{T}}$ (and sometimes by $\mathbf{A}'$), is the matrix obtained from $\mathbf{A}$ by interchanging rows and columns to produce the $n \times m$ matrix

$$\mathbf{A}^{\mathrm{T}} = [a_{ij}]^{\mathrm{T}} = [a_{ji}].$$

The definition of the transpose of a matrix means that the first *row* of $\mathbf{A}$ becomes the first *column* of $\mathbf{A}^{\mathrm{T}}$, the second *row* of $\mathbf{A}$ becomes the second *column* of $\mathbf{A}^{\mathrm{T}}$, ... , and, finally, the mth *row* of $\mathbf{A}$ becomes the mth *column* of $\mathbf{A}^{\mathrm{T}}$. In particular, if $\mathbf{A}$ is a row vector, then its transpose is a column vector, and conversely.

EXAMPLE 3.2

$$\text{If } \mathbf{A} = \begin{bmatrix} 2 & 6 & 3 \\ 1 & 0 & 4 \end{bmatrix} \quad \text{then} \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 2 & 1 \\ 6 & 0 \\ 3 & 4 \end{bmatrix}, \quad \text{and if } \mathbf{A} = [7,\ 3,\ 2] \quad \text{then } \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 7 \\ 3 \\ 2 \end{bmatrix}.$$

■

Definition of scaling a matrix by a number

Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix and λ be a scalar (real or complex). Then if $\mathbf{A}$ is **scaled** by λ, written $\lambda\mathbf{A}$, every element of $\mathbf{A}$ is multiplied by λ to yield the $m \times n$ matrix

$$\lambda\mathbf{A} = [\lambda a_{ij}].$$

EXAMPLE 3.3

$$\text{If } \lambda = 2 \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 2 & -6 & 7 \\ 1 & 4 & 15 \end{bmatrix}, \quad \text{then } \lambda\mathbf{A} = 2\mathbf{A} = \begin{bmatrix} 4 & -12 & 14 \\ 2 & 8 & 30 \end{bmatrix},$$

and if $\lambda = -1$, then

$$\lambda\mathbf{A} = (-1)\mathbf{A} = -\mathbf{A} = \begin{bmatrix} -2 & 6 & -7 \\ -1 & -4 & -15 \end{bmatrix}.$$

■

difference (subtraction) of matrices

Taken together, the definitions of the addition and scaling of matrices show that if the matrices $\mathbf{A}$ and $\mathbf{B}$ are conformable for addition, then the subtraction of matrix $\mathbf{B}$ from $\mathbf{A}$, called their **difference** and written $\mathbf{A} - \mathbf{B}$, is to be interpreted as

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}.$$

EXAMPLE 3.4

$$\text{If } \mathbf{A} = \begin{bmatrix} 2 & 5 & 8 \\ 1 & -4 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 4 & 5 \\ 2 & -4 & 1 \end{bmatrix}, \quad \text{then } \mathbf{A} - \mathbf{B} = \begin{bmatrix} 0 & 1 & 3 \\ -1 & 0 & 4 \end{bmatrix}.$$

■

negative of a matrix

The **null** or **zero** matrix $\mathbf{0}$ is defined as any matrix in which every element is zero. The introduction of the null matrix makes it appropriate to call $-\mathbf{A}$ the **negative** of $\mathbf{A}$, because

$$\mathbf{A} - \mathbf{A} = \mathbf{A} + (-1)\mathbf{A} = \mathbf{0}.$$

When working with the null matrix the number of its rows and columns is never stated, because these are always taken to be whatever is appropriate for the equation that is involved.

Definition of the product of a row and a column vector

Let $\mathbf{a} = [a_1, a_2, \ldots, a_r]$ be an r-element row vector, and $\mathbf{b} = [b_1, b_2, \ldots, b_r]^{\mathrm{T}}$ be an r-element column vector. Then the product $\mathbf{ab}$, in this order, is the number defined as

$$\mathbf{ab} = a_1b_1 + a_2b_2 \cdots + a_rb_r.$$

Notice that this product is *only* defined when the number of elements in the row vector $\mathbf{A}$ equals the number of elements in the column vector $\mathbf{B}$.

EXAMPLE 3.5

Find the product $\mathbf{ab}$ given that $\mathbf{a} = [1,\ 4,\ -3,\ 10]$ and $\mathbf{b} = [2,\ 1,\ 4,\ -2]^{\mathrm{T}}$.

Solution

$$\begin{aligned}\mathbf{ab} &= [1,\ 4,\ -3,\ 10]\begin{bmatrix} 2 \\ 1 \\ 4 \\ -2 \end{bmatrix} \\ &= (1)\cdot(2) + (4)\cdot(1) + (-3)\cdot(4) + (10)\cdot(-2) \\ &= -26.\end{aligned}$$

Definition of the product of matrices

Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix in which the rth row is the row vector $\mathbf{a}_r$, and let $\mathbf{B} = [b_{ij}]$ be a $p \times q$ matrix in which the sth column is the column vector $\mathbf{b}_s$. The matrix product $\mathbf{AB}$, in this order, is only defined if the number of columns in $\mathbf{A}$ equals the number of rows in $\mathbf{B}$, so that $n = p$. The product is then an $m \times q$ matrix with the element in its rth row and sth column defined as $\mathbf{a}_r\mathbf{b}_s$. Thus, if $c_{rs} = \mathbf{a}_r\mathbf{b}_s$, as $c_{rs} = a_{r1}b_{1s} + a_{r2}b_{2s} + \cdots + a_{rn}b_{ns}$,

$$\mathbf{AB} = [c_{rs}] = [a_{r1}b_{1s} + a_{r2}b_{2s} + \cdots + a_{rn}b_{ns}],$$

for $1 \le r \le m$ and $1 \le s \le q$, or, equivalently,

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1\mathbf{b}_1 & \mathbf{a}_1\mathbf{b}_2 & \mathbf{a}_1\mathbf{b}_3 & \ldots & \mathbf{a}_1\mathbf{b}_q \\ \mathbf{a}_2\mathbf{b}_1 & \mathbf{a}_2\mathbf{b}_2 & \mathbf{a}_2\mathbf{b}_3 & \ldots & \mathbf{a}_2\mathbf{b}_q \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{a}_m\mathbf{b}_1 & \mathbf{a}_m\mathbf{b}_2 & \mathbf{a}_m\mathbf{b}_3 & \ldots & \mathbf{a}_m\mathbf{b}_q \end{bmatrix}.$$

When a matrix product $\mathbf{AB}$ is defined, the matrices are said to be **conformable** for matrix multiplication in the given order.

---

in general, matrix multiplication is noncommutative

It is important to notice that when the product $\mathbf{AB}$ is defined, the product $\mathbf{BA}$ may or may not be defined, and even when $\mathbf{BA}$ is defined, in general $\mathbf{AB} \neq \mathbf{BA}$. This situation is recognized by saying that, in general, matrix multiplication is **noncommutative**.

Provided matrices $\mathbf{A}$ and $\mathbf{B}$ are conformable for multiplication, the above rule for finding their product $\mathbf{AB}$, in this order, is best remembered by saying that the element in the ith row and jth column of $\mathbf{AB}$ is the product of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$.

EXAMPLE 3.6

Form the matrix products $\mathbf{AB}$ and $\mathbf{BA}$ given that

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & -3 \\ 2 & 5 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 4 & 1 \\ 2 & 6 \\ 0 & 3 \end{bmatrix}.$$

Solution Let us calculate the matrix product $\mathbf{AB}$. The first and second row vectors of $\mathbf{A}$ are $\mathbf{a}_1 = [1, 4, -3]$ and $\mathbf{a}_2 = [2, 5, 4]$, and the first and second column vectors of $\mathbf{B}$ are $\mathbf{b}_1 = [4, 2, 0]^{\mathrm{T}}$ and $\mathbf{b}_2 = [1, 6, 3]^{\mathrm{T}}$. As $\mathbf{A}$ is a 2×3 matrix and $\mathbf{B}$ is a 3×2 matrix, the product $\mathbf{AB}$ is conformable for multiplication and yields a 2×2 matrix

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1\mathbf{b}_1 & \mathbf{a}_1\mathbf{b}_2 \\ \mathbf{a}_2\mathbf{b}_1 & \mathbf{a}_2\mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} (1\cdot4 + 4\cdot2 + (-3)\cdot0) & (1\cdot1 + 4\cdot6 + (-3)\cdot3) \\ (2\cdot4 + 5\cdot2 + 4\cdot0) & (2\cdot1 + 5\cdot6 + 4\cdot3) \end{bmatrix}$$
$$= \begin{bmatrix} 12 & 16 \\ 18 & 44 \end{bmatrix}.$$

The product $\mathbf{BA}$ is also conformable for multiplication and yields a 3×3 matrix, where now we must use the *row* vectors of $\mathbf{B}$ that with an obvious change of notation are $\mathbf{b}_1 = [4, 1]$, $\mathbf{b}_2 = [2, 6]$, $\mathbf{b}_3 = [0, 3]$, and the *column* vectors of $\mathbf{A}$ that are $\mathbf{a}_1 = [1, 2]^{\mathrm{T}}$, $\mathbf{a}_2 = [4, 5]^{\mathrm{T}}$, and $\mathbf{a}_3 = [-3, 4]^{\mathrm{T}}$, so that

$$\mathbf{BA} = \begin{bmatrix} \mathbf{b}_1\mathbf{a}_1 & \mathbf{b}_1\mathbf{a}_2 & \mathbf{b}_1\mathbf{a}_3 \\ \mathbf{b}_2\mathbf{a}_1 & \mathbf{b}_2\mathbf{a}_2 & \mathbf{b}_2\mathbf{a}_3 \\ \mathbf{b}_3\mathbf{a}_1 & \mathbf{b}_3\mathbf{a}_2 & \mathbf{b}_3\mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} (4\cdot1 + 1\cdot2) & (4\cdot4 + 1\cdot5) & (4\cdot(-3) + 1\cdot4) \\ (2\cdot1 + 6\cdot2) & (2\cdot4 + 6\cdot5) & (2\cdot(-3) + 6\cdot4) \\ (0\cdot1 + 3\cdot2) & (0\cdot4 + 3\cdot5) & (0\cdot(-3) + 3\cdot4) \end{bmatrix}$$
$$= \begin{bmatrix} 6 & 21 & -8 \\ 14 & 38 & 18 \\ 6 & 15 & 12 \end{bmatrix}.$$

This is an example of two matrices $\mathbf{A}$ and $\mathbf{B}$ that can be combined to form the products $\mathbf{AB}$ and $\mathbf{BA}$, but $\mathbf{AB} \neq \mathbf{BA}$. ■

EXAMPLE 3.7

Write the system of simultaneous equations (1) in matrix form.

Solution Using the matrices $\mathbf{A}$ and $\mathbf{b}$ in (2) and setting $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ allows the system of equations (1) to be written

$$\mathbf{Ax} = \mathbf{b}.$$

Here, as is usual, to save space the transpose operation has been used to display the elements of column vector $\mathbf{x}$ in the more convenient form $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$. ■

The definitions of matrix multiplication and addition lead almost immediately to the results of the following theorem, so the proof is left as an exercise.

THEOREM 3.1

some important properties of matrices

Associative and distributive properties of matrices Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be matrices that are conformable for the operations that follow, and let λ be a scalar. Then:

(i) If $\mathbf{AB}$ and $\mathbf{BA}$ are both defined, in general $\mathbf{AB} \neq \mathbf{BA}$;

(ii) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} = \mathbf{ABC}$;

(iii) $(\lambda\mathbf{A})\mathbf{B} = \mathbf{A}(\lambda\mathbf{B}) = \lambda\mathbf{AB}$;

(iv) $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$;

(v) $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$. ■

THEOREM 3.2

Transposition of a product If matrices $\mathbf{A}$ and $\mathbf{B}$ are conformable to form the product $\mathbf{AB}$, then

$$(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}.$$

Proof The products $(\mathbf{AB})^{\mathrm{T}}$ and $\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$ are both defined, and each is an $m \times q$ matrix. Introduce the notation $[\mathbf{M}]_{ij}$ to denote the element of $\mathbf{M}$ in row i and column j. Then from the transpose operation and the rule for matrix multiplication, for all permissible i, j,

$$[\mathbf{AB}]^{\mathrm{T}}_{i,j} = [\mathbf{AB}]_{j,i} = (\text{product of } j\text{th row of } \mathbf{A} \text{ with } i\text{th column of } \mathbf{B}) = \sum_{k=1}^{n} a_{jk}b_{ki}.$$

Similarly,

$$\begin{aligned}[\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}]_{i,j} &= (\text{product of } i\text{th row of } \mathbf{B}^{\mathrm{T}} \text{ with } j\text{th column of } \mathbf{A}^{\mathrm{T}}) \\ &= (\text{product of } i\text{th column of } \mathbf{B} \text{ with } j\text{th row of } \mathbf{A}) = \sum_{k=1}^{n} a_{jk}b_{ki}.\end{aligned}$$

So $[\mathbf{AB}]^{\mathrm{T}}_{ij} = [\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}]_{ij}$ for all permissible i, j, showing that $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$. ■

raising a matrix to a power

It is an immediate consequence of Theorem 3.1(ii) that if $\mathbf{A}$ is a square matrix and m and n are positive integers,

$$\underbrace{\mathbf{A}\cdot\mathbf{A}\cdot\mathbf{A}\cdot\ \ldots\ \cdot\mathbf{A}}_{n \text{ times}} = \mathbf{A}^{n} \quad \text{and} \quad \mathbf{A}^{m}\cdot\mathbf{A}^{n} = \mathbf{A}^{m+n}.$$

A useful result from the definition of addition is

$$(\mathbf{A}+\mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}+\mathbf{B}^{\mathrm{T}},$$

while from Theorem 3.2

$$(\mathbf{ABC})^{\mathrm{T}} = \mathbf{C}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}.$$

As the order in which a sequence of permissible matrix multiplications is performed influences the product, it is necessary to introduce a form of words that makes the order unambiguous. This is accomplished by saying that if matrix **A** multiplies matrix **B** from the *left*, as in **AB**, then **B** is **premultiplied** by **A**, while if **A** multiplies **B** from the *right*, as in **BA**, then **B** is **postmultiplied** by **A**. Equivalently, in the product **AB**, we can say that **A** is *postmultiplied* by **B**, or that **B** is *premultiplied* by **A**.

pre- and post-multiplication of matrices

Important Differences Between Ordinary Algebraic Equations and Matrix Equations

(i) The algebraic equation $ab = 0$, in which a and b are numbers, not both of which are zero, implies that either $a = 0$ or $b = 0$. However, if the matrix product **AB** is defined and is such that $\mathbf{AB} = \mathbf{0}$, then it does *not* necessarily follow that either $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.

(ii) The algebraic equation $ab = ac$ in which a, b, and c are numbers, with $a \neq 0$, allows cancellation of the factor a leading to the conclusion that $b = c$. However, if the matrix products **AB** and **AC** are defined and are such that $\mathbf{AB} = \mathbf{AC}$, this does *not* necessarily imply that $\mathbf{B} = \mathbf{C}$, so that cancellation of matrix factors is *not* permissible.

The validity of these two statements can be seen by considering the following simple examples.

EXAMPLE 3.8

Consider matrices **A** and **B** given by

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 3 & 12 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 4 & -8 \\ -1 & 2 \end{bmatrix}.$$

Then $\mathbf{AB} = \mathbf{0}$, but neither **A** nor **B** is a null matrix. ■

EXAMPLE 3.9

Consider the matrices **A**, **B**, and **C** given by

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 4 & 6 \\ 2 & 3 & 4 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 3 & 6 & 8 \\ 3 & 5 & 6 \end{bmatrix}.$$

Then

$$\mathbf{AB} = \mathbf{AC} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix},$$

but $\mathbf{B} \neq \mathbf{C}$. ■

In a square $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, the elements on a line extending from top left to bottom right is called the **leading diagonal** of **A**, and it contains the n elements $a_{11}, a_{22}, \ldots, a_{nn}$.

So the leading diagonal of the 2×2 matrix **A** in Example 3.8 contains the elements 1 and 12, and the leading diagonal of the 2×2 matrix **B** contains the elements 4 and 2. Symbolically, the leading diagonal of the $n \times n$ matrix $\mathbf{A} = [a_{ij}]$ shown below comprises the n elements in the shaded diagonal strip, though these

leading diagonal and trace of a matrix

n elements do *not* form an n element vector.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdot & \cdot & \cdot & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdot & \cdot & \cdot & a_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix}.$$

The **trace** of a square matrix $\mathbf{A}$, written $\mathrm{tr}(\mathbf{A})$, is the sum of the terms on its leading diagonal, so for the foregoing matrix $\mathrm{tr}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn}$.

Square matrices in which all elements away from the leading diagonal are zero, but not every element on the leading diagonal is zero, are called **diagonal matrices**. Of the class of diagonal matrices, the most important are the **unit** matrices, also called **identity** matrices, in which every element on the leading diagonal is the number 1. These $n \times n$ matrices are usually all denoted by the symbol $\mathbf{I}$, with the value of n being understood to be appropriate to the context in which they arise. If, however, the value of n needs to be indicated, the symbol $\mathbf{I}$ can be replaced by $\mathbf{I}_n$. It is easily seen from the definition of matrix multiplication that for any $m \times n$ matrix $\mathbf{A}$ it follows that

$$\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n \text{ or, more simply, } \mathbf{IA} = \mathbf{AI} = \mathbf{A},$$

and that when $\mathbf{A}$ is an $n \times n$ matrix,

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}.$$

When working with matrices, the unit matrix $\mathbf{I}$ plays the part of the unit real number, and it is because of this that $\mathbf{I}$ is called either the *unit* or the *identity* matrix.

identity or unit matrix

An example of a 4×4 diagonal matrix is

$$\mathbf{D} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \text{with the trace given by } \mathrm{tr}(\mathbf{D}) = 3 + 2 + 0 + 1 = 6.$$

The 3×3 unit matrix is the diagonal matrix

$$\mathbf{I} = \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{and its trace is } \mathrm{tr}(\mathbf{I}) = 1 + 1 + 1 = 3.$$

Various special square $n \times n$ matrices occur sufficiently frequently for them to be given names, and some of the most important of these are the following:

some special matrices

Upper triangular matrices are matrices in which all elements below the leading diagonal are zero. A typical example of a 4×4 upper triangular matrix is

$$\mathbf{U} = \begin{bmatrix} 1 & 3 & -1 & 0 \\ 0 & 2 & -6 & 1 \\ 0 & 0 & -3 & 2 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Lower triangular matrices are matrices in which all elements above the leading diagonal are zero. A typical example of a 4×4 lower triangular matrix is

$$\mathbf{L} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 3 & -2 & 5 & 0 \\ -2 & 4 & 7 & 3 \end{bmatrix}.$$

Symmetric matrices $\mathbf{A} = [a_{ij}]$ are matrices in which $a_{ij} = a_{ji}$ for all i and j. If $\mathbf{A}$ is symmetric, then $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$. A typical example of a symmetric matrix is

$$\mathbf{M} = \begin{bmatrix} 1 & 5 & -3 \\ 5 & 4 & 2 \\ -3 & 2 & 7 \end{bmatrix}.$$

Skew-symmetric matrices $\mathbf{A} = [a_{ji}]$ are matrices in which $a_{ij} = -a_{ji}$ for all i and j. From the definition of an $n \times n$ skew-symmetric matrix we have $a_{ii} = -a_{ii}$ for $i = 1, 2, \ldots, n$, so the elements on the leading diagonal must all be zero. An equivalent definition of a skew-symmetric matrix $\mathbf{A}$ is that $\mathbf{A}^{\mathrm{T}} = -\mathbf{A}$. A typical example of a skew-symmetric matrix is

$$\mathbf{S} = \begin{bmatrix} 0 & 3 & -5 & 6 \\ -3 & 0 & 2 & -4 \\ 5 & -2 & 0 & -1 \\ -6 & 4 & 1 & 0 \end{bmatrix}.$$

An **orthogonal** matrix $\mathbf{Q}$ is a matrix such that $\mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}$. A typical orthogonal matrix is

$$\mathbf{Q} = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \end{bmatrix}.$$

More special than the preceding real valued matrices are matrices $\mathbf{A} = [a_{ij}]$ in which the elements a_{ij} are complex numbers. We will write $\overline{\mathbf{A}}$ to denote the matrix obtained from $\mathbf{A}$ by replacing each of its elements a_{ij} by its complex conjugate $\overline{a}_{ij}$, so that

$$\overline{\mathbf{A}} = [\overline{a}_{ij}].$$

Then matrix $\mathbf{A}$ is said to be **Hermitian** if

$$\overline{\mathbf{A}}^{\mathrm{T}} = \mathbf{A}.$$

A typical Hermitian matrix is

$$\mathbf{A} = \begin{bmatrix} 7 & 1 - 4i \\ 1 + 4i & 3 \end{bmatrix}.$$

The matrix $\mathbf{A}$ is said to be **skew-Hermitian** if

$$\overline{\mathbf{A}}^{\mathrm{T}} = -\mathbf{A}.$$

A typical skew-Hermitian matrix is

$$\mathbf{A} = \begin{bmatrix} 3i & 5 + 2i \\ -5 + 2i & 0 \end{bmatrix}.$$

More will be said later about some of these special square matrices and the ways in which they arise.

block matrices

Finally, we mention that every $m \times n$ matrix $\mathbf{A}$ can be represented differently as a **block matrix**, in which each element is itself a matrix. This is accomplished by **partitioning** the matrix $\mathbf{A}$ into **submatrices** by considering horizontal and vertical lines to be drawn through $\mathbf{A}$ between some of its rows and columns, and then identifying each group of elements so defined as a **submatrix** of $\mathbf{A}$. Clearly there is more than one way in which a matrix can be partitioned. As an example of matrix partitioning, let us consider the 3×3 matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 2 & 0 \\ 2 & 1 & 0 \end{bmatrix}.$$

One way in which this matrix can be partitioned is as follows:

$$\mathbf{A} = \left[\begin{array}{cc:c} 3 & -1 & 2 \\ \hdashline 1 & 2 & 0 \\ 2 & 1 & 0 \end{array}\right].$$

This can now be written in block matrix form as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

where the submatrices are

$$\mathbf{A}_{11} = [3 \ \ -1], \quad \mathbf{A}_{12} = [2], \quad \mathbf{A}_{21} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{A}_{22} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The addition and scaling of block matrices follow the same rules as those for ordinary matrices, but care must be exercised when multiplying block matrices. To see how multiplication of block matrices can be performed, let us consider the product of matrix $\mathbf{A}$ above and the 3×4 matrix

$$\mathbf{B} = \left[\begin{array}{c:ccc} 1 & 2 & 2 & 1 \\ 3 & 1 & 1 & 0 \\ \hdashline 2 & 3 & 0 & 2 \end{array}\right],$$

which are conformable for the product $\mathbf{AB}$ that is itself a 3×4 matrix. If $\mathbf{B}$ is partitioned as indicated by the dashed lines, it can be written as

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix},$$

where the submatrices are

$$\mathbf{B}_{11} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \mathbf{B}_{12} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{B}_{21} = [2], \quad \text{and} \quad \mathbf{B}_{22} = [3, \ 0, \ 2].$$

Consideration of the definition of the product of matrices shows that we may now write the matrix product $\mathbf{AB}$ in the condensed form

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix},$$

where the partitioned matrices have been multiplied as though their elements were ordinary elements. This result follows because of correct partitioning, as each product of submatrices is conformable for multiplication and all of the matrix sums are conformable for addition.

In this illustration, routine calculations show that

$$\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} = [4], \quad \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = [11,\ 5,\ 7],$$
$$\mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}, \quad \text{and} \quad \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \begin{bmatrix} 4 & 4 & 1 \\ 5 & 5 & 2 \end{bmatrix},$$

so

$$\mathbf{AB} = \begin{bmatrix} [4] & [11,\ 5,\ 7] \\ \begin{bmatrix} 7 \\ 5 \end{bmatrix} & \begin{bmatrix} 4 & 4 & 1 \\ 5 & 5 & 2 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 4 & 11 & 5 & 7 \\ 7 & 4 & 4 & 1 \\ 5 & 5 & 5 & 2 \end{bmatrix}.$$

This result is easily confirmed by direct matrix multiplication.

The calculation of a matrix product **AB** using partitioned matrices applies in general, provided the partitioning of **A** and **B** is performed in such a way that the products of all the submatrices involved are defined.

Matrix partitioning has various uses, one of which arises in machine computation when a very large fixed matrix **A** needs to be multiplied by a sequence of very large matrices **P**, **Q**, **R**, ... If it happens that **A** can be partitioned in such a way that some of its submatrices are null matrices, the computational time involved can be drastically reduced, because the product of a submatrix and a null matrix is a null matrix, and so need not be computed. The economy follows from the fact that in machine computation multiplications occupy most of the time, so any reduction in their number produces a significant reduction in the time taken to evaluate a matrix product, and the result is even more significant when the same partitioned matrix with null blocks is involved in a sequence of calculations.

Block matrices are also of significance when describing complex oscillation problems governed by a large system of simultaneous ordinary differential equations. Their importance arises from the fact that the matrix of coefficients of the equations often contains many null submatrices, and when this happens the structure of the nonnull blocks provides useful information about the fundamental modes of oscillation that are possible, and also about their interconnections.

For other accounts of elementary matrices see the appropriate chapters in references [2.1], [2.5], and [2.7] to [2.12].

Summary

This section defined $m \times n$ matrices, and the special cases of column and row vectors, and it introduced the fundamental algebraic operations of equality, addition, scaling, transposition, and multiplication of matrices. It was shown that, in general, matrix multiplication is not commutative, so that even when both of the products **AB** and **BA** are defined, it is usually the case that $\mathbf{AB} \neq \mathbf{BA}$.

Pre- and postmultiplication of matrices was defined, and some important special types of matrices were introduced, such as the unit matrix **I**. It was also shown how a matrix **A** can be subdivided into blocks, and that a matrix operation performed on **A** can be interpreted in terms of matrix operations performed on block matrices obtained by subdivision of **A**.

EXERCISES 3.1

In Exercises 1 through 4 find the values of the constants a, b, and c in order that $\mathbf{A} = \mathbf{B}$.

1. $\mathbf{A} = \begin{bmatrix} a^2 & 1 & c \\ 2 & 3 & a \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} -a & 1 & 4 \\ 2 & b & -1 \end{bmatrix}$.

2. $\mathbf{A} = \begin{bmatrix} 1 & 4 & 3 \\ a & 2 & 4 \\ 9 & 1 & c \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 2 & 4 \\ b & 1 & 0 \end{bmatrix}$.

3. $\mathbf{A} = \begin{bmatrix} a^2 & a & 1 \\ b & 1 & 2 \\ 1+a & 2+c & 6 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} a^2 & a & 1 \\ 3 & 1 & 2 \\ 2 & 4 & 6 \end{bmatrix}$.

4. $\mathbf{A} = \begin{bmatrix} 1 & 3+a & 2 \\ 1+b & a & 5 \\ b^2 & 1 & a^2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & -1 & c \\ 4 & a & 5 \\ b^2 & 1 & a^2 \end{bmatrix}$.

In Exercises 5 through 8 find $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$.

5. $\mathbf{A} = \begin{bmatrix} 1 & 4 & 3 & 6 \\ 2 & 1 & 0 & 2 \\ 1 & -1 & 0 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & 0 & 1 & -2 \\ 1 & 1 & -3 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$.

6. $\mathbf{A} = \begin{bmatrix} 1 & 7 & 6 \\ 0 & 2 & 4 \\ -1 & 0 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & -1 & 6 \\ 1 & -2 & 3 \\ 2 & 1 & 2 \end{bmatrix}$.

7. $\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 1 & 0 \\ 1 & 1 & 0 \\ 2 & 2 & 4 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 2 & 3 \\ 3 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 3 & 2 \end{bmatrix}$.

8. $\mathbf{A} = \begin{bmatrix} 1 & 4 & 3 & 6 \\ 0 & 2 & 1 & 4 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 \\ 1 & 1 & 1 & 3 \end{bmatrix}$.

In Exercises 9 through 12 form the sum $\lambda\mathbf{A} + \mu\mathbf{B}$.

9. $\lambda = 1$, $\mu = 3$, $\mathbf{A} = \begin{bmatrix} 1 & 4 & 2 \\ 2 & 1 & 4 \\ 3 & 2 & 2 \end{bmatrix}$,

$\mathbf{B} = \begin{bmatrix} 2 & 3 & -1 \\ 1 & 2 & 4 \\ 1 & 0 & 3 \end{bmatrix}$.

10. $\lambda = -1$, $\mu = 2$, $\mathbf{A} = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 4 & 0 \end{bmatrix}$,

$\mathbf{B} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix}$.

11. $\lambda = 4$, $\mu = -2$, $\mathbf{A} = \begin{bmatrix} 4 & 3 & 1 \\ 2 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$,

$\mathbf{B} = \begin{bmatrix} 6 & 1 & 0 \\ 2 & 4 & 2 \\ 1 & 1 & 2 \end{bmatrix}$.

12. $\lambda = 3$, $\mu = -3$, $\mathbf{A} = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 2 & 1 \\ 3 & 6 & 2 \end{bmatrix}$,

$\mathbf{B} = \begin{bmatrix} 3 & 2 & 1 \\ 4 & 2 & 3 \\ 2 & 1 & 1 \end{bmatrix}$.

In Exercises 13 through 16 find the product $\mathbf{AB}$.

13. $\mathbf{A} = [1,\ 4,\ -2,\ 3]$, $\mathbf{B} = [2,\ 1,\ -1,\ 2]^{\mathrm{T}}$.

14. $\mathbf{A} = [2,\ 3,\ 1,\ 4]$, $\mathbf{B} = [3,\ 1,\ 1,\ 3]^{\mathrm{T}}$.

15. $\mathbf{A} = [1,\ 4,\ 3,\ 7,\ 5]$,
$\mathbf{B} = [2,\ 2,\ -1,\ -1,\ 3]^{\mathrm{T}}$.

16. $\mathbf{A} = [1,\ 3,\ -1,\ 2,\ 0]$,
$\mathbf{B} = [-1,\ 2,\ 13,\ 4,\ 1]^{\mathrm{T}}$.

In Exercises 17 through 22 find the product $\mathbf{AB}$ and, when it exists, the product $\mathbf{BA}$.

17. $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 1 \end{bmatrix}$.

18. $\mathbf{A} = [1,\ 4,\ 6,\ -7]$, $\mathbf{B} = [2,\ 3,\ -2,\ 3]^{\mathrm{T}}$.

19. $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 1 & -5 \\ 7 & 2 & 0 \end{bmatrix}$.

20. $\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 9 & -1 & 4 \\ 1 & 6 & -2 \\ 2 & 2 & 3 \end{bmatrix}$.

21. $\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 1 & 2 \\ 2 & 2 & 6 \\ 1 & 5 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 5 & 2 & 3 \\ 2 & 0 & 4 \\ 1 & 4 & 7 \end{bmatrix}$.

22. $\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 1 & 1 & 4 \\ 1 & 0 & 2 & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 6 & -2 \\ -1 & 4 \end{bmatrix}$.

23. Given

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & -3 \\ 5 & 1 & 4 \\ -3 & 4 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 5 & 6 \\ 1 & 6 & 3 \end{bmatrix},$$

show that $(\mathbf{AB})^{\mathrm{T}} = \mathbf{BA}$.

In Exercises 24 through 28 write the given systems of equations in the matrix form $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is the coefficient matrix, $\mathbf{x}$ is the vector of unknowns, and $\mathbf{b}$ is the nonhomogeneous vector term.

24. $3x + 5y - 6z = 7$
$x - 7y + 4z = -3$
$2x + 4y - 5z = 4.$

25. $4u + 5v - w + 7z = 25$
$3u + 2v + 3z = 6$
$v + 6w - 7z = 0.$

26. $5x + 3y - 6z = 14$
$6x - 5y + 11z = 20$
$x - 4y + 3z = 2$
$9x - 3y + 2z = 35.$

27. $3x + 4y - 2z = \lambda x$
$2x - 7y + 6z = \lambda y$
$8x + 3y + 5z = \lambda z.$

28. $2x + 3y + 6z = \lambda(3x + 2y + 3z)$
$3x - 4y + 2z = \lambda(x - 5y + 2z)$
$4x + 9y + 2z = \lambda(x - 2y + 4z).$

29. If

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 6 \\ 1 & 2 & 0 \\ 0 & 1 & 3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 0 & 1 \\ 4 & 2 & 3 \\ 0 & -1 & 1 \end{bmatrix}, \quad \text{and}$$

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix},$$

solve for $\mathbf{X}$ given that

$$3\mathbf{X} + \mathbf{A} = \mathbf{A}^{\mathrm{T}}\mathbf{B} - \mathbf{X} + 3\mathbf{B}.$$

30. If

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 2 & 1 \\ 3 & 0 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}, \quad \text{and}$$

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix},$$

solve for $\mathbf{X}$ given that

$$2\mathbf{AB}^{\mathrm{T}} + \mathbf{X} - 2\mathbf{I} = 3\mathbf{X} + 4\mathbf{B} - 2\mathbf{A}.$$

31. Given that

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 4 \end{bmatrix},$$

show that

$$\mathbf{A}^3 - 9\mathbf{A}^2 + 18\mathbf{A} = \mathbf{0}.$$

32. Given that

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 1 & -2 \end{bmatrix},$$

show that

$$\mathbf{A}^3 + 2\mathbf{A}^2 - \mathbf{A} - 2\mathbf{I} = \mathbf{0}.$$

33. Prove the second result in Theorem 3.1 that $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} = \mathbf{ABC}$.

34. Prove the third result in Theorem 3.1 that $(\lambda\mathbf{A})\mathbf{B} = \mathbf{A}(\lambda\mathbf{B}) = \lambda\mathbf{AB}$.

35. Prove the fourth result in Theorem 3.1 that $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

In Exercises 36 through 39 verify that $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\,\mathbf{A}^{\mathrm{T}}$.

36. $\mathbf{A} = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 5 \\ 0 & 2 & 1 \end{bmatrix}.$

37. $\mathbf{A} = \begin{bmatrix} 2 & 1 & 4 & 3 \\ 1 & 6 & 2 & 1 \\ 1 & 1 & -2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 1 & 5 \\ -1 & 3 & 2 \\ 1 & 7 & 3 \end{bmatrix}.$

38. $\mathbf{A} = \begin{bmatrix} 1 & 4 & 2 \\ 7 & 3 & -1 \\ 0 & 2 & 5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & 1 & -5 \\ 1 & 3 & 4 \\ 2 & 0 & 8 \end{bmatrix}.$

39. $\mathbf{A} = \begin{bmatrix} 1 & 4 & 6 & 2 \\ 2 & 1 & 4 & 1 \\ 3 & 0 & 0 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & 1 & 4 \\ 2 & 2 & 5 \\ 1 & 1 & 1 \end{bmatrix}.$

40. Verify that $(\mathbf{ABC})^{\mathrm{T}} = \mathbf{C}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$ given that

$$\mathbf{A} = \begin{bmatrix} 1 & 5 \\ 3 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & -2 \\ 4 & 5 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} -2 & 3 \\ 5 & 7 \end{bmatrix}.$$

41. Prove that if $\mathbf{D}$ is the $n \times n$ diagonal matrix

$$\mathbf{D} = \begin{bmatrix} k_1 & 0 & 0 & \cdots & 0 \\ 0 & k_2 & 0 & \cdots & 0 \\ 0 & 0 & k_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & k_n \end{bmatrix}, \quad \text{then}$$

$$\mathbf{D}^m = \begin{bmatrix} k_1^m & 0 & 0 & \cdots & 0 \\ 0 & k_2^m & 0 & \cdots & 0 \\ 0 & 0 & k_3^m & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & k_n^m \end{bmatrix},$$

where m is a positive integer.

42. Find $\mathbf{A}^2$, $\mathbf{A}^3$, and $\mathbf{A}^4$, given that

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 7 \\ 2 & 5 & 6 \\ 1 & 0 & -1 \end{bmatrix}.$$

43. Find $\mathbf{A}^2$, $\mathbf{A}^4$, and $\mathbf{A}^6$, given that

$$\mathbf{A} = \begin{bmatrix} 1/2 & -(\sqrt{3})/2 \\ (\sqrt{3})/2 & 1/2 \end{bmatrix}.$$

44. Use the matrix $\mathbf{A}$ in Exercise 42 to find $\mathbf{A}^3$, $\mathbf{A}^5$, and $\mathbf{A}^7$.

45. A square matrix $\mathbf{A}$ such that $\mathbf{A}^2 = \mathbf{A}$ is said to be **idempotent**. Find the three idempotent matrices of the form

$$\mathbf{A} = \begin{bmatrix} 1 & p \\ q & r \end{bmatrix}.$$

46. A square matrix $\mathbf{A}$ such that for some positive integer n has the property that $\mathbf{A}^{n-1} \neq \mathbf{0}$, but $\mathbf{A}^n = \mathbf{0}$ is said to be **nilpotent of index** $\boldsymbol{n}$ ($n \geq 2$). Show that the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 4 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$$

is nilpotent and find its index.

47. A **quadratic form** in the variables $x_1, x_2, x_3, \ldots, x_n$ is an expression of the form $ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + \cdots + fx_{n-1}x_n + gx_n^2$ in which some of the coefficients $a, b, c, d, \ldots, f, g$ may be zero. Explain why $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ is a quadratic form and find the quadratic form for which

$$\mathbf{A} = \begin{bmatrix} 3 & 4 & 0 & 3 \\ 4 & 2 & 2 & 6 \\ 0 & 2 & 5 & 1 \\ 3 & 6 & 1 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

48. Find the quadratic form $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ when

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 3 & 6 \\ 2 & 3 & 5 & 4 \\ 1 & 4 & 1 & 2 \\ 2 & 0 & 4 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

49. Explain why the matrix $\mathbf{A}$ in the general expression for a quadratic form $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ can always be written as a *symmetric* matrix.

In Exercises 50 through 52 find the symmetric matrix $\mathbf{A}$ for the given quadratic form when written $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, with $\mathbf{x} = [x,\ y,\ z]^{\mathrm{T}}$.

50. $x^2 + 3xy - 4y^2 + 4xz + 6yz - z^2$.

51. $2x^2 + 4xy + 6y^2 + 7xz - 9z^2$.

52. $7x^2 + 7xy - 5y^2 + 4xz + 2yz - 9z^2$.

53. A square matrix $\mathbf{P}$ is called a **stochastic matrix** if all its elements are nonnegative and the sum of the elements in each row is 1. Thus, the matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

will be a stochastic matrix if $p_{ij} \geq 0$ for $0 \leq i \leq n$, $0 \leq j \leq n$, and

$$\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for } i = 1, 2, \ldots, n.$$

Let the n element column vector $\mathbf{E} = [1, 1, 1, \ldots, 1]^{\mathrm{T}}$. By considering the matrix product $\mathbf{PE}$, and using mathematical induction, prove that $\mathbf{P}^m$ is a stochastic matrix for all positive integral values of m.

54. Construct a 3×3 stochastic matrix $\mathbf{P}$. Find $\mathbf{P}^2$ and $\mathbf{P}^3$, and by showing that all elements of these matrices are nonnegative and that all their row-sums are 1, verify the result of Exercise 53 that each of these matrices is a stochastic matrix.

3.2 Some Problems That Give Rise to Matrices

(a) Electric Circuits with Resistors and Applied Voltages

A simple electric circuit involving five resistors and three applied voltages is shown in Fig. 3.1. The directions of the currents i_1, i_2, and i_3 flowing in each branch of the circuit are shown by arrows. The currents themselves can be determined by an application of *Ohm's* law and the *Kirchhoff* laws that can be stated as follows:

(a) Voltage = current × resistance (Ohm's law);

(b) The algebraic sum of the potential drops around each closed circuit is zero (Kirchhoff's second law);

(c) The current entering each junction must equal the algebraic sum of the currents leaving it (Kirchhoff's first law).

equations and matrices for electric circuits

An application of these laws to the circuit in Fig. 3.1, where the potentials are in volts, the resistances are in ohms, and the currents are in amps, leads to the following

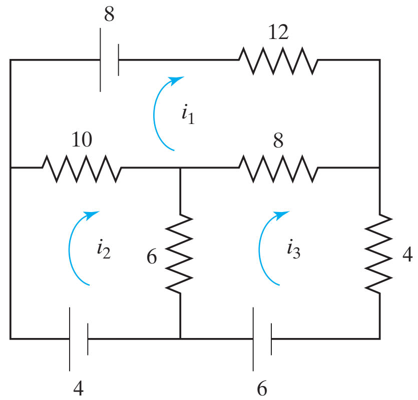


FIGURE 3.1 An electric circuit with resistors and applied voltages.

set of simultaneous equations:

$$8 = 12i_1 + 10(i_1 - i_2) + 8(i_1 - i_3)$$
$$4 = 10(i_2 - i_1) + 6(i_2 - i_3)$$
$$6 = 8(i_3 - i_1) + 6(i_3 - i_2) + 4i_3.$$

After collecting terms this system can be written as the matrix equation $\mathbf{Ax} = \mathbf{b}$, with

$$\mathbf{A} = \begin{bmatrix} 30 & -10 & -8 \\ -10 & 16 & -6 \\ -8 & -6 & 18 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 8 \\ 4 \\ 6 \end{bmatrix}.$$

The directions assumed for the currents i_r for $r = 1, 2, 3$ are shown by the arrows in Fig. 3.1, but if after the system of equations is solved, the value of the current is found to be negative, the direction of its arrow must be reversed.

The circuit in Fig. 3.1 is simple, so in this example the currents can be found by routine elimination between the three equations. When many coupled circuits are involved a matrix approach is useful, and it then becomes necessary to develop a method for solving for $\mathbf{x}$ the matrix equation $\mathbf{Ax} = \mathbf{b}$, the elements of which are the required currents. If the number of equations is small, $\mathbf{x}$ can be found by making use of the matrix $\mathbf{A}^{-1}$, inverse to $\mathbf{A}$, that will be introduced later, though the most computationally efficient approach is to use one of the numerical methods for solving systems of linear simultaneous equations described in Chapter 19.

(b) Combinatorial Problems: Graph Theory

Matrices play an important role in combinatorial problems of many different types and, in particular, in graph theory. The purpose of the brief account offered here will be to illustrate a particular application of matrices, and no attempt will be made to discuss their subsequent use in the solution of the associated problems.

Combinatorial problems involve dealing with the possible arrangements of situations of various different kinds, and computing the number and properties of such arrangements. The arrangements may be of very diverse types, involving at one extreme the ordering of matches that are to take place in a tennis tournament,

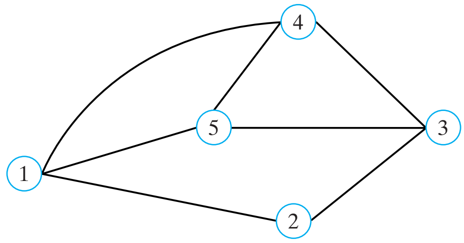


FIGURE 3.2 The graph representing routes.

and at the other extreme finding an optimum route for a delivery truck or for the most efficient routing of work through a machine shop.

The ideas involved are most easily illustrated by means of examples, the first of which involves the delivery from a storage depot of a consumable product to a group of supermarkets in a large city where it is important that daily deliveries be made as rapidly as possible. One possibility involves a delivery truck making a delivery to each supermarket in turn and returning to the storage depot between each delivery before setting out on the next delivery.

An alternative is to travel between supermarkets after each delivery without returning to the storage depot. The question that then arises is which approach to routing is the best, and how it is to be determined.

A typical situation is illustrated in Fig. 3.2, in which supermarkets numbered 1 to 5 are involved, with circles representing supermarkets and lines and arcs representing the routes.

graphs, vertices, edges, and adjacency matrix

The representation in Fig. 3.2 is called a **graph**, and it is to be regarded as a set of points represented by the circles called **vertices** of the graph, and **edges** of the graph represented by the lines and arcs. In Fig. 3.2 the vertices are the circles 1, 2, . . . , 5 and the seven edges are the lines and arcs connecting the vertices.

A special type of matrix associated with such a graph is an **adjacency matrix**, that is, a matrix whose only entries are 0 or 1. The rules for the entries in an adjacency matrix $\mathbf{A} = [a_{ij}]$ are that

$$a_{ij} = \begin{cases} 1, & \text{if vertices } i \text{ and } j \text{ are joined by an edge} \\ 0, & \text{otherwise.} \end{cases}$$

The adjacency matrix for the graph in Fig. 3.2 is seen to be the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

It is to be expected that an adjacency matrix is symmetric, because if i is adjacent to j, then j is adjacent to i.

Although we shall not attempt to do so here, the interconnection properties of the problem represented by the graph in Fig. 3.2 can be analyzed in terms of its adjacency matrix $\mathbf{A}$. The optimum routing problem can then be resolved once the traveling times along roads (lines or arcs) are known.

Sometimes it happens that the edges in a graph represent connections that only operate in one direction, so then arrows must be added to the graph to indicate these

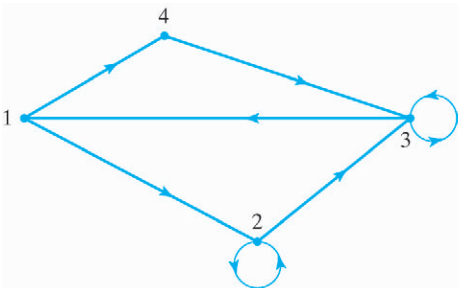


FIGURE 3.3 A typical digraph.

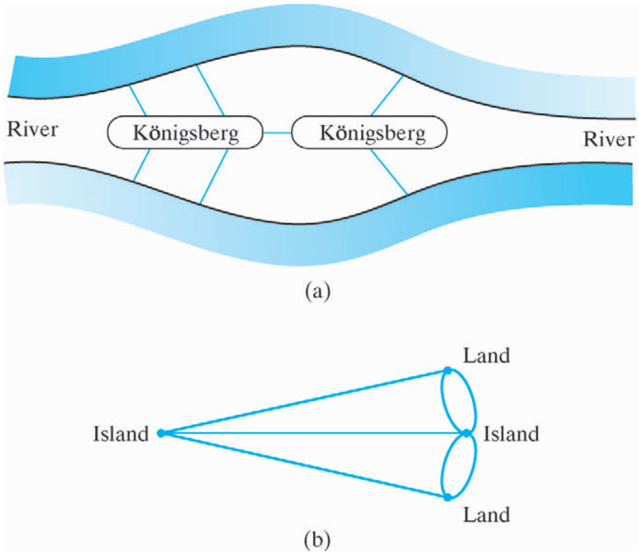


FIGURE 3.4 The Königsberg bridge problem.

digraph

directions. A graph of this type is called a **digraph** (directed graph). The rules for the entries in the adjacency matrix $\mathbf{A} = [a_{ij}]$ of a digraph are that

$$a_{ij} = \begin{cases} 1, & \text{if vertices } i \text{ and } j \text{ are joined by an edge with an arrow from } i \text{ to } j \\ 0, & \text{otherwise.} \end{cases}$$

A typical digraph is shown in Fig. 3.3, and it has the associated adjacency matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The adjacency matrix **A** characterizes all the possible interconnections between the four vertices and, as with the previous example, an analysis of the properties of any situation capable of representation in terms of this digraph can be performed using the matrix **A**. Problems of this type can arise in transportation problems in cities with one-way streets, and in chemical processes where a fluid is piped to different parts of a plant through an interconnecting network of pipes through which fluid may only flow in a given direction.

Königsberg bridge problem

Before closing this brief introduction to graph theory, mention should be made of a problem of historical significance, since it represented the start of graph theory as it is known today. The problem is called the **Königsberg bridge problem**, and it was solved by Euler (1707–1783). During the early 18th century the Prussian town of Königsberg was established on two adjacent islands in the middle of the river Pregel. The islands were linked to the land on either side of the river, and to one another, by seven bridges, as shown in Fig. 3.4a. It was suggested to Euler that he should resolve the conjecture that it ought to be possible to walk through the town, starting and ending at the same place, while crossing each of the seven bridges only once.

Euler replaced the picture in Fig. 3.4a by the graph in Fig. 3.4b, though it was not until much later that the term *graph* in the sense used here was introduced. In Fig. 3.4b the vertices S and Q represent the two islands and, using the same lettering, P and R represent the riverbanks. The number of edges incident on each vertex represents the number of bridges connected to the corresponding land mass. Euler introduced the concept of a *connected graph*, in which each pair of vertices is linked by a set of edges, and also what is now called an *eulerian circuit*, comprising a path through all vertices that starts and ends at the same vertex and uses every edge only once. He called the number of edges incident upon a vertex the *degree* of the vertex, and by using these ideas he was able to prove the impossibility of the conjecture. The arguments involved are not difficult, but their details would be out of place here.

Many more practical problems are capable of solution by graph theory, which itself belongs to the branch of mathematics called *combinatorics*. In elementary accounts, graph theory and related combinatorial issues are usually called *discrete mathematics*. More information about combinatorics and its connection with matrices can be found in References [2.2] and [2.13].

(c) Translations, Rotations, and Scaling of Graphs: Computer Graphics

matrices and computer graphics

The simplest operations in computer graphics involve copying a picture to a different location, rotating a picture about a fixed point, and scaling a picture, where the scaling can be different in the horizontal and vertical directions. These operations are called, respectively, a **translation**, a **rotation**, and a **scaling** of the picture. Operations of this nature can all be represented in terms of matrices, and they involve what are called **linear transformations** of the original picture.

Translation

A translation of a two-dimensional picture involves copying it to a different location without either rotating it or changing its horizontal and vertical scales. Figure 3.5 shows the original cartesian axes $O(x, y)$ and the shifted axes $O'(x', y')$, where the respective axes remain parallel to their original directions and the origin O' is located at the point (h, k) relative to the $O(x, y)$ axes.

The relationship between the two sets of coordinates is given by

$$x = x' + h \quad \text{and} \quad y = y' + k.$$

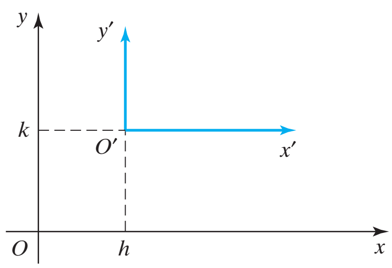


FIGURE 3.5 A translation.

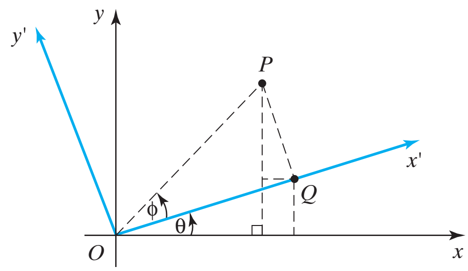


FIGURE 3.6 A rotation through an angle θ.

If $\mathbf{x} = [x, y]^{\mathrm{T}}$, $\mathbf{x}' = [x', y']^{\mathrm{T}}$, and $\mathbf{b} = [h, k]^{\mathrm{T}}$, the coordinate transformation can be written in matrix form as

$$\mathbf{x} = \mathbf{x}' + \mathbf{b},$$

where matrix $\mathbf{b}$ represents the translation.

Rotation

A rotation of the coordinate axes through an angle θ is shown in Fig. 3.6, where $P(x, y)$ is an arbitrary point. The coordinates of P in the (x, y) reference frame and the (x', y') reference frame are related as

$$\begin{aligned} x &= OR = OP\cos(\phi + \theta) = OP\cos\phi\cos\theta - OP\sin\phi\sin\theta \\ &= OQ\cos\theta - PQ\sin\theta = x'\cos\theta - y'\sin\theta, \end{aligned}$$

and

$$\begin{aligned} y &= PR = OP\sin(\phi + \theta) = OP\sin\phi\cos\theta + OP\cos\phi\sin\theta \\ &= PQ\cos\theta + OQ\sin\theta = y'\cos\theta + x'\sin\theta, \end{aligned}$$

so

$$x = x'\cos\theta - y'\sin\theta \quad \text{and} \quad y = y'\cos\theta + x'\sin\theta.$$

Defining the matrices $\mathbf{x}$, $\mathbf{x}'$, and $\mathbf{R}$ as

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix}, \quad \text{and} \quad \mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

allows the coordinate transformation to be written as

$$\mathbf{x} = \mathbf{R}\mathbf{x}'.$$

Scaling

If $\mathbf{S}$ is a matrix of the form

$$\mathbf{S} = \begin{bmatrix} k_x & 0 \\ 0 & k_y \end{bmatrix},$$

where k_x and k_y are positive constants, and $\mathbf{x}' = \mathbf{S}\mathbf{x}$, it follows that

$$x = k_x x' \quad \text{and} \quad y = k_y y',$$

showing that x is obtained by scaling x' by k_x, while y is obtained by scaling y' by k_y. This form of scaling is represented by premultiplication of $\mathbf{x}$ by $\mathbf{S}$, and if, for example,

$$\mathbf{S} = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix},$$

the effect of this transformation on a circle of radius a will be to map it into an ellipse with semimajor axis of length $4a$ parallel to the x-axis and a semiminor axis of length $3a$ parallel to the y-axis.

Composite transformations

By combining the preceding matrix operations to form a **composite transformation**, it is possible to carry out several transformations simultaneously. As an example, the effect of a rotation $\mathbf{R}$ followed by a translation $\mathbf{b}$ when performed on a vector $\mathbf{x}'$ are seen to be described by the matrix equation

$$\mathbf{x} = \mathbf{R}\mathbf{x}' + \mathbf{b},$$

the effect of which is shown in Fig. 3.7.

If a scaling $\mathbf{S}$ is performed before the rotation and translation, the effect on a vector $\mathbf{x}'$ is described by the matrix equation

$$\mathbf{x} = \mathbf{R}\mathbf{S}\mathbf{x}' + \mathbf{b}.$$

This is illustrated in Fig. 3.8b, which shows the effect when a transformation of this type is performed on the circle of radius a centered on the origin shown in Fig. 3.8a, with

$$\mathbf{b} = \begin{bmatrix} h \\ k \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \cos \pi/3 & -\sin \pi/3 \\ \sin \pi/3 & \cos \pi/3 \end{bmatrix}, \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}.$$

It is seen that the circle has first been scaled to become an ellipse with semiaxes $3a$ and $2a$, after which the ellipse has been rotated through an angle $\pi/3$, and finally its center has been translated to the point (h, k).

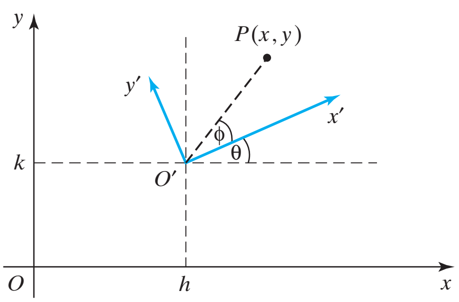


FIGURE 3.7 A rotation and a translation.

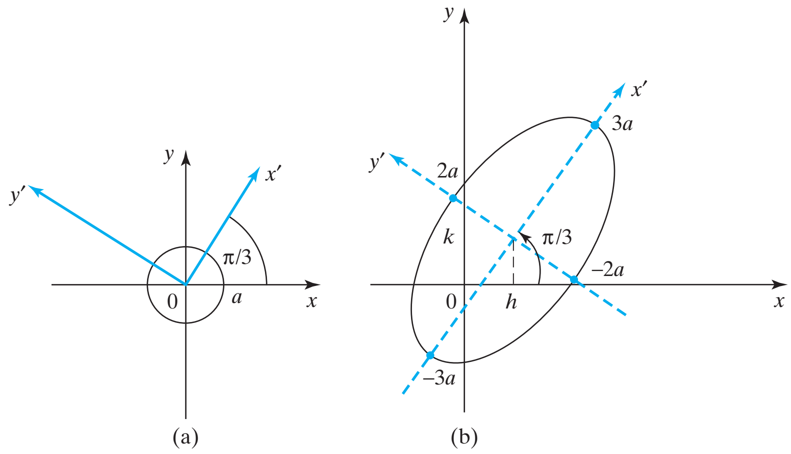


FIGURE 3.8 The composite transformation $\mathbf{x} = \mathbf{RSx}' + \mathbf{b}$.

It is essential to remember that the *order* in which transformations are performed will, in general, influence the result. This is easily seen by considering the two transformations $\mathbf{x} = \mathbf{RSx}' + \mathbf{b}$ and $\mathbf{x} = \mathbf{SRx}' + \mathbf{b}$. If the first of these is performed on the circle in Fig. 3.8a, it produces Fig. 3.8b, but when the second is performed on the same circle, it first converts it into an ellipse with its major axis horizontal, and then translates the center of the ellipse to the point (h, k). In this case the effect of the rotation cannot be seen, because the circle is symmetric with respect to rotations.

A relationship of the form $\mathbf{x} = \mathbf{F}(\mathbf{x}')$ can be interpreted geometrically in two distinct ways which are equally valid:

1. As the change in the way we describe the location of a point P. Then the relationship is called a transformation of coordinates (Figs. 3.5, 3.6, 3.7).
2. As a mapping of a point P from one location to a new one.

(d) Matrix Analysis of Framed Structures

A **framed structure** is a network of straight struts joined at their ends to form a rigid three-dimensional structure. A typical framed structure is the steel work for a large building before the walls and floors have been added. A simple example of a framed structure, called a **truss**, is a plane construction in which the struts are joined together to form triangles, as in the side section of the small bridge shown in Fig. 3.9.

For safety, to ensure that no strut fails when the bridge carries the largest permitted load, it is necessary to determine the force experienced by each strut in the truss when the bridge supports its maximum load in several different positions. Typically, the largest load could be due to a heavy truck crossing the bridge. The analysis of trusses is usually simplified by making the following assumptions:

- The structure is in the vertical plane;
- The weight of each strut can be neglected;
- Struts are rigid and so remain straight;

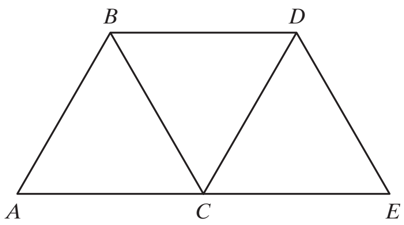


FIGURE 3.9 A typical truss found in a side section of a bridge.

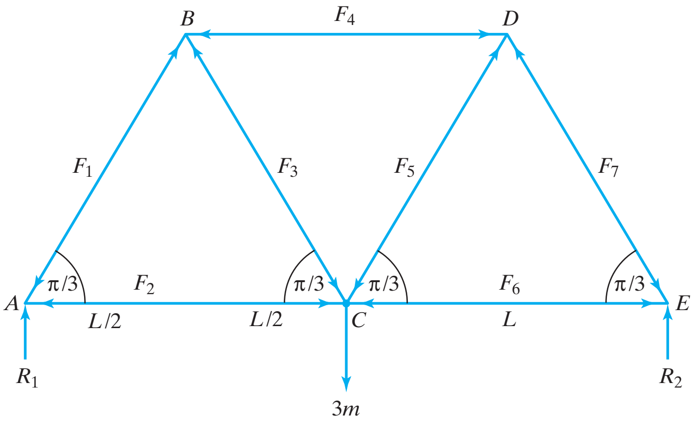


FIGURE 3.10 A truss supporting a concentrated load.

- Each joint is considered to be hinged, so the only forces acting at a joint are along the struts meeting at the joint if forces are applied at joints only.
- There are no redundant struts, so that removing a strut will cause the truss to collapse.

We now write down the simultaneous equations that must be solved to find the forces acting in the seven struts of length L that form the truss shown in Fig. 3.10, when a concentrated load $3m$ is located at point C midway between A and E. This load could be considered to be a heavily laden truck standing in the center of the bridge.

To determine the reactions at the support points A and E, we use the fact that for equilibrium the turning moments about these two points must be zero. The turning moment of the load $3m$ about the point A must be cancelled by the turning moment of the reaction R_2 at E, so $3m(L) = R_2(2L)$, showing that $R_2 = 3m/2$. Similarly, the turning moment of the load $3m$ about the point E must be cancelled by the turning moment of the reaction R_1 at A, so $3m(L) = R_1(2L)$, showing that $R_1 = 3m/2$.

The directions in which the forces F_1 to F_7 are assumed to act are shown by arrows, and if later a force is found to be negative, the direction of the associated arrows must be reversed. For equilibrium the sum of the vertical components of all forces acting at each joint must be zero, as must be the sum of the horizontal components of all forces acting at each joint. The equations representing the balance of forces at each joint are as follows, where when resolving the forces acting at joint C, the effect of the load $3m$ which acts vertically downwards must be taken into account:

equations and matrices for a framed structure

Joint A(vertical)	$F_1 \sin \pi/3 - 3m/2 = 0$
Joint A(horizontal)	$F_1 \cos \pi/3 + F_2 = 0$
Joint B(vertical)	$F_1 \sin \pi/3 + F_3 \sin \pi/3 = 0$
Joint B(horizontal)	$F_1 \cos \pi/3 - F_3 \cos \pi/3 - F_4 = 0$
Joint C(vertical)	$F_3 \sin \pi/3 + F_5 \sin \pi/3 + 3m = 0$
Joint C(horizontal)	$F_2 + F_3 \cos \pi/3 - F_5 \cos \pi/3 - F_6 = 0$
Joint D(vertical)	$F_5 \sin \pi/3 + F_7 \sin \pi/3 = 0$

Joint D(horizontal) $\quad F_4 + F_5 \cos \pi/3 - F_7 \cos \pi/3 = 0$
Joint E(vertical) $\quad F_7 \sin \pi/3 - 3m/2 = 0$
Joint E(horizontal) $\quad F_6 + F_7 \cos \pi/3 = 0.$

After substituting for $\sin \pi/3$ and $\cos \pi/3$, these equations can be written in the matrix form $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2}\sqrt{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2}\sqrt{3} & 0 & \frac{1}{2}\sqrt{3} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} & -1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\sqrt{3} & 0 & \frac{1}{2}\sqrt{3} & 0 & 0 \\ 0 & 1 & \frac{1}{2} & 0 & -\frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2}\sqrt{3} & 0 & \frac{1}{2}\sqrt{3} \\ 0 & 0 & 0 & 1 & \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}\sqrt{3} \\ 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{2} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \\ F_7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3m/2 \\ 0 \\ 0 \\ 0 \\ -3m \\ 0 \\ 0 \\ 0 \\ 3m/2 \\ 0 \end{bmatrix}.$$

These are 10 equations for the 7 unknown forces F_1 to F_7, so unless 3 of the equations represented in $\mathbf{Ax} = \mathbf{b}$ are combinations of the remaining 7 equations, we cannot expect there to be a solution. When the **rank** of a matrix is introduced in Section 3.6, we will see how systems of this type can be checked for consistency and, when appropriate, simplified and solved.

In this case the equations are sufficiently simple that they can be solved sequentially, without the use of matrices. The solution is seen to be

$$F_1 = m\sqrt{3}, \quad F_2 = -m/(\sqrt{3}/2), \quad F_3 = -m\sqrt{3}, \quad F_4 = m\sqrt{3},$$
$$F_5 = -m\sqrt{3}, \quad F_6 = -m(\sqrt{3}/2), \quad F_7 = m\sqrt{3}.$$

The signs show that the arrows in Fig. 3.10 associated with forces F_2, F_3, F_5, and F_6 should be reversed, so these struts are in tension, while the others are in compression.

Notice that matrix **A** is determined by the geometry of the truss, and so does not change when forces are applied to more than one of the joints on the truss (bridge). This means that after the 10 equations have been reduced to seven, the same modified matrix **A** can be use to find the forces in the struts for *any* form of concentrated loading. Had a more complicated struss been involved, many more equations would have been involved, so that a matrix approach becomes necessary. This approach also identifies any redundant struts in a structure, because the force in a redundant strut is indeterminate.

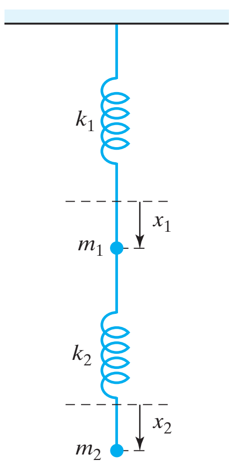


FIGURE 3.11 A compound mass–spring system.

(e) A Compound Mass–Spring System

Matrices can have variables as elements, and an analysis of the compound mass–spring system shown in Fig. 3.11 shows one way in which this can arise. Figure 3.11 represents a mass m_1 suspended from a rigid support by a spring of negligible mass with spring constant k_1, and a mass m_2 suspended from mass m_1 by a spring of negligible mass with spring constant k_2. The vertical displacement of m_1 from its

equilibrium position is x_1, and the vertical displacement of m_2 from its equilibrium position is x_2. Each spring is assumed to be linearly elastic, so the restoring force exerted by a spring is equal to the product of the displacement from its equilibrium position and the spring constant.

The product of the mass m_1 and its acceleration is $m_1 d^2x_1/dt^2$, and the restoring force due to spring k_1 is k_1x_1, while the restoring force due to spring k_2 is $k_2(x_1 - x_2)$, so the equation of motion of m_1 is

equations of motion of a coupled mass–spring system

$$m_1\frac{d^2x_1}{dt^2} = -k_1x_1 - k_2(x_1 - x_2).$$

Similarly, the equation of motion of m_2 is

$$m_2\frac{d^2x_2}{dt^2} = -k_2(x_2 - x_1),$$

where the negative signs are necessary because the springs act to restore the masses to their original positions.

This system can be written as the matrix differential equation $\ddot{\mathbf{x}} + \mathbf{A}\mathbf{x} = \mathbf{0}$, by defining $\mathbf{A}$ and $\mathbf{x}$ as

$$\mathbf{A} = \begin{bmatrix} \dfrac{(k_1 + k_2)}{m_1} & -\dfrac{k_2}{m_1} \\ -\dfrac{k_2}{m_2} & \dfrac{k_2}{m_2} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \text{and} \quad \ddot{\mathbf{x}} = \begin{bmatrix} \dfrac{d^2x_1}{dt^2} \\ \dfrac{d^2x_2}{dt^2} \end{bmatrix}.$$

The solution of this system will not be considered here as ordinary differential equations and systems of the type derived here are discussed in detail in Chapter 6, where matrix methods are also developed. Chapter 7 develops Laplace transform methods for the solution of differential equations and systems. It will suffice to mention here that the dynamical behavior of the compound mass–spring system in Fig. 3.11 is completely characterized by matrix $\mathbf{A}$.

(f) Stochastic Processes

Certain problems arise that are not of a deterministic nature, so that both the formulation of the problem and its outcome must be expressed in terms of probabilities. The probability p that a certain event occurs is a number in the interval $0 \le p \le 1$. An event with probability $p = 0$ is one that never occurs, and an event with probability $p = 1$ is one that is certain to occur. So, for example, when tossing a coin N times and recording each outcome as an H (head) or a T (tail), if the number of heads is N_{H} and the number of tails is N_{T}, so that $N = N_{\mathrm{H}} + N_{\mathrm{T}}$, the numbers N_{H}/N and N_{T}/N will be approximations to the respective probabilities that a head or a tail occurs when the coin is tossed. If the coin is *unbiased*, it is reasonable to expect that as N increases both N_{H}/N and N_{T}/N will approach the value 1/2. This will mean, of course, that the chances of either a head or a tail occurring on each occasion are equal.

The example we now outline is called a **stochastic process** and is illustrated by considering a process that evolves with time and is such that at any given moment it may be in precisely one of N different situations, usually called **states**, where N is finite. We shall denote the N states in which the process may find itself at any given time t_m by $S_1, S_2, \ldots, S_N$, with $m = 0, 1, 2, \ldots,$ and $t_{m-1} < t_m$, it being assumed that the outcome at each time depends on probabilities, and so is *not* deterministic.

To formulate the problem we assume that what are called the **conditional probabilities** p_{ki} (also called **transition probabilities**) that determine the probability with which the process will be in state S_j at time t_m are all known, given that it was in state S_k at time t_{m-1}, and that these probabilities are the same from t_1 to t_2 as from t_{m-1} to t_m for $m = 0, 1, 2, \ldots$. This last assumption means that the probability with which the transition from state S_k to S_j occurs is *independent* of the time at which the process was in state S_k.

The conditional probabilities can be arranged as the $N \times N$ matrix $\mathbf{P} = [p_{jk}]$, so as probabilities are involved, all the p_{jk} are nonnegative, and as each stage must have an outcome, the sum of the elements in every row of matrix $\mathbf{P}$ must equal 1. A matrix $\mathbf{P}$ with these properties, namely that

$$0 \le p_{jk} \le 1, \quad 0 \le j \le N, \quad 0 \le k \le N, \quad \text{and} \quad \sum_{j=1}^{N} p_{jk} = 1,$$

is called a **stochastic matrix** (see Exercise 53, Section 3.1).

stochastic matrix and a Markov process

Processes like these, whose condition at any subsequent instant does not depend on how the process arrived at its present state, are called **Markov processes**. Simple but typical examples of such processes involving only two states are gambling wins and losses, the reliability of machines that may either be operational or under repair, shells fired from a gun that either hit or miss the target and errors that introduce an incorrect digit 1 or 0 when transferring binary coded information.

To develop the argument a little further, let us now consider a process that can be in one of two states, and that the matrix $\mathbf{P}$ describing its transitions is given by

$$\mathbf{P} = \begin{bmatrix} 2/3 & 1/3 \\ 1/4 & 3/4 \end{bmatrix}.$$

Now suppose that initially the probability distribution is given by the row matrix $\mathbf{E}(0) = [p, q]$, where, of course, $p + q = 1$. Then if $\mathbf{E}(m)$ denotes the probability distribution of the states at time t_m, it follows that $\mathbf{E}(1) = \mathbf{E}(0)\mathbf{P}$, but as $\mathbf{P}$ is independent of the time we conclude that after m transitions the general result must be

$$\mathbf{E}(m) = \mathbf{E}(0)\mathbf{P}^m,$$

so in this case

$$\mathbf{E}(m) = [p, q] \begin{bmatrix} 2/3 & 1/3 \\ 1/4 & 3/4 \end{bmatrix}^m.$$

Direct calculation shows that

$$\mathbf{E}(3) = [0.470p + 0.398q, 0.530p + 0.602q],$$
$$\mathbf{E}(6) = [0.432p + 0.426q, 0.568p + 0.574q],$$

and

$$\mathbf{E}(10) = [0.429p + 0.429q, 0.571p + 0.571q],$$

so it is reasonable to ask if $\mathbf{E}(m)$ tends to a limiting vector as $m \to \infty$ and, if so, what this is? As this problem is simple, an analytical answer is possible, though it involves using a *diagonalizing* matrix $\mathbf{P}$ which will be discussed later.

We will see later that $\mathbf{P}$ can be written as $\mathbf{ADB}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{AB} = \mathbf{I}$. In this case

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 5/12 \end{bmatrix}, \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix},$$

so

$$\mathbf{P} = \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5/12 \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix}.$$

In what follows we will need to make repeated use of the fact that

$$\mathbf{BA} = \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$$

Using this last property we find that

$$\begin{aligned} \mathbf{P}^2 &= \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5/12 \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5/12 \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5/12 \end{bmatrix}^2 \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix}. \end{aligned}$$

However, when a diagonal matrix is raised to a power, each of its elements is raised to that same power (see Problem 41, Section 3.1), so

$$\mathbf{P}^2 = \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (5/12)^2 \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix}$$

and, in general,

$$\mathbf{P}^m = \begin{bmatrix} 1 & 4 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (5/12)^m \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 1/7 & -1/7 \end{bmatrix}.$$

Thus,

$$\mathbf{P}^m = \begin{bmatrix} \dfrac{3 + 4(5/12)^m}{7} & \dfrac{4 - 4(5/12)^m}{7} \\ \dfrac{3 - 3(5/12)^m}{7} & \dfrac{4 + 3(5/12)^m}{7} \end{bmatrix},$$

showing that as $m \to \infty$, so

$$\lim_{m\to\infty} \mathbf{E}(m)\mathbf{P}^m = [3(p+q)/7, 4(p+q)/7] = [3/7, 4/7],$$

and we have found the limiting state of the system.

Stochastic processes also occur that involve more than two states. The problem of determining the probability with which such processes will be in a given state, and when a limiting state exists, the limiting values of the probabilities involved, is of considerable practical importance. An introduction to stochastic process can be found in reference [2.4].

Summary

This section has introduced some of the many areas in which matrices play an essential role. These range from electric circuits needing the application of Kirchhoff's laws, through routing problems involving the concepts of directed graphs and adjacency matrices, to the classical Königsberg bridge problem, computer graphic operations performed by linear transformations, the matrix analysis of forces in a framed structure, the oscillations of a coupled mass–spring system, and stochastic processes.

EXERCISES 3.2

1. State which of the following matrices is a stochastic matrix, giving a reason when this is not the case.

(a) $\begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.25 & 0 & 0.75 \\ 0.5 & 0.5 & 0 \end{bmatrix}$. (c) $\begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.7 & 0.3 & 0.2 \\ 0.4 & 0.2 & 0.4 \end{bmatrix}$.

(b) $\begin{bmatrix} 1.2 & 0 & -0.2 \\ 0 & 0.8 & 0.2 \\ 0.6 & 0.3 & 0.1 \end{bmatrix}$. (d) $\begin{bmatrix} 0.3 & 0.1 & 0.6 \\ 0.8 & 0 & 0.2 \\ 0 & 1 & 0 \end{bmatrix}$.

2. Given the stochastic matrix

$$\mathbf{P} = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{bmatrix}$$

and the initial probability distribution $\mathbf{E}(0) = [p, q]$, with $p, q \geq 0$ and $p + q = 1$, the probability distribution of the two states at time t_m is given by

$$\mathbf{E}(m) = \mathbf{E}(0)\mathbf{P}^m.$$

Find $\mathbf{E}(2)$, $\mathbf{E}(4)$, and $\mathbf{E}(6)$, together with their values when $p = 1/4$, $q = 3/4$.

In Exercises 3 through 6 find the adjacency matrices for the given graphs and digraphs.

3.

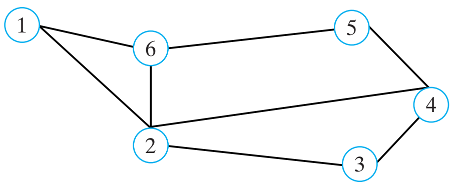


FIGURE 3.12

4.

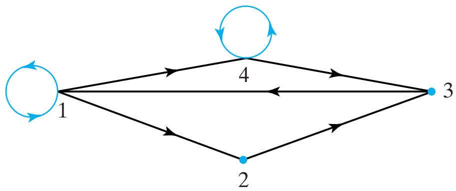


FIGURE 3.13

5.

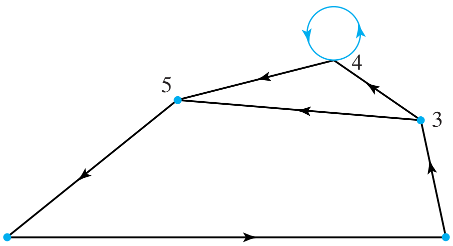


FIGURE 3.14

6.

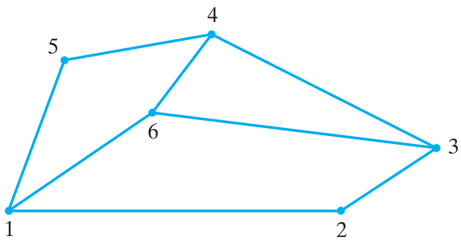


FIGURE 3.15

3.3 Determinants

Every square matrix **A** with numbers as elements has associated with it a single unique number called the *determinant* of **A**, which is written det**A**. If **A** is an $n \times n$ matrix, the determinant of **A** is indicated by displaying the elements a_{ij} of **A** between two vertical bars as follows:

notation for a determinant

$$\det\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{5}$$

The number n is called the **order** of determinant **A**, and in (5) the vertical bars are used to distinguish det**A**, that is a number, from matrix **A** that is an $n \times n$ array of numbers.

A general definition of the value of det$\mathbf{A}$ in terms of its elements a_{ij} will be given later, so for the moment we define only the value of first and second order determinants (see Section 1.7). If $\mathbf{A}$ only contains a single element a_{11} so $\mathbf{A} = [a_{11}]$ then, by definition, det$\mathbf{A} = a_{11}$, and if $\mathbf{A}$ is the 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

then, by definition,

$$\det \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{21}a_{12}. \tag{6}$$

Notice that in (6) the numerical value of det$\mathbf{A}$ is obtained by forming the product of the two terms a_{11} and a_{22} on the leading diagonal, and subtracting from it the product of the two terms a_{21} and a_{12} on the cross diagonal. This process, called **expanding** the determinant, is easily remembered by representing the method by which the determinant is expanded as

$$\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} = a_{11}a_{22} - a_{21}a_{12},$$

where the product involving the downward arrow generates the first pair of terms on the right and the product involving the upward arrow indicates that the product of the associated pair of terms is to be subtracted.

EXAMPLE 3.10 Find det$\mathbf{A}$ given

(a) det$\mathbf{A} = \begin{vmatrix} 3 & -1 \\ 2 & 6 \end{vmatrix}$ and (b) det$\mathbf{A} = \begin{vmatrix} 1+i & i \\ -3i & 2 \end{vmatrix}$.

Solution **(a)** Using (5) we have

$$\det\mathbf{A} = \begin{vmatrix} 3 & -1 \\ 2 & 6 \end{vmatrix} = 3 \cdot 6 - 2 \cdot (-1) = 20.$$

(b) Again using (5) we have

$$\det\mathbf{A} = \begin{vmatrix} 1+i & i \\ -3i & 2 \end{vmatrix} = (1+i) \cdot 2 - (-3i) \cdot i = -1 + 2i. \quad \blacksquare$$

To provide some motivation for the introduction of determinants, we solve by elimination the two linear simultaneous algebraic equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2. \end{aligned} \tag{7}$$

To eliminate x_2 we multiply the first equation by a_{22} and the second equation by a_{12}, and then subtract the results to obtain

$$(a_{11}a_{22} - a_{21}a_{12})x_1 = a_{22}b_1 - a_{12}b_2.$$

This shows that when $a_{11}a_{22} - a_{21}a_{12} \neq 0$,

$$x_1 = \frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{21}a_{12}}.$$

This result can be expressed in terms of det$\mathbf{A}$ as

$$x_1 = (a_{22}b_1 - a_{12}b_2)/\det\mathbf{A}. \tag{8}$$

Similarly, when x_1 is eliminated from equations (7) we find that

$$x_2 = (a_{11}b_2 - a_{21}b_1)/\det\mathbf{A}. \tag{9}$$

Cramer's rule for a system of two equations

Examination of (8) and (9) shows that their numerators can be written in terms of determinants that are closely related to det$\mathbf{A}$, because

$$x_1 = \frac{D_1}{D} \quad \text{and} \quad x_2 = \frac{D_2}{D}, \tag{10}$$

where

$$D = \det\mathbf{A}, \quad D_1 = \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}, \quad \text{and} \quad D_2 = \begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}. \tag{11}$$

The form of solution of equations (7) in terms of the determinants in (10) and (11) is called **Cramer's rule**. The rule itself says that $x_i = D_i/D$ for $i = 1, 2$, where determinant D_1 is obtained from $D = \det\mathbf{A}$ by replacing the *first* column of $\mathbf{A}$ by the nonhomogeneous terms b_1 and b_2 on the right of equations (7), and determinant D_2 is obtained from D by replacing the *second* column of $\mathbf{A}$ by these same two terms.

EXAMPLE 3.11

Use Cramer's rule to solve the equations

$$3x_1 + 5x_2 = 4$$
$$2x_1 - 4x_2 = 1.$$

Solution The three determinants required by Cramer's are

$$D = \det\mathbf{A} = \begin{vmatrix} 3 & 5 \\ 2 & -4 \end{vmatrix} = -22, \quad D_1 = \begin{vmatrix} 4 & 5 \\ 1 & -4 \end{vmatrix} = -21, \quad D_2 = \begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix} = -5,$$

so $x_1 = D_1/D = 21/22$ and $x_2 = D_2/D = 5/22$. ■

This example shows how determinants enter naturally into the solution of a system of equations. As determinants of order $n > 2$ occur in the study of differential equations, analytical geometry, throughout linear algebra, and elsewhere, it is necessary to generalize the definition of a determinant of order 2 given in (6) to determinants of any order n.

With this objective in mind, we first define the *minors* and *cofactors* of a determinant of order n. The **minor** M_{ij} associated with the element a_{ij} in the ith row and jth column of the nth order determinant in (5) is the determinant of order $n - 1$ formed from det$\mathbf{A}$ by deleting the elements in the ith row and jth column. As each element of det$\mathbf{A}$ has an associated minor, a determinant of order n has n^2 minors.

minors and cofactors

By way of example, the minor M_{3j} of the nth order determinant in (5) is the determinant of order $n - 1$

$$M_{3j} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1j-1} & a_{1j+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j-1} & a_{2j+1} & \cdots & a_{2n} \\ a_{41} & a_{42} & \cdots & a_{4j-1} & a_{4j+1} & \cdots & a_{4n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nj-1} & a_{nj+1} & \cdots & a_{nn} \end{vmatrix}. \tag{12}$$

The **cofactor** C_{ij} associated with the element a_{ij} in determinant (5) is defined in terms of the minor M_{ij} as

$$C_{ij} = (-1)^{i+j} M_{ij} \quad \text{for } i, j = 1, 2, \ldots, n, \tag{13}$$

so an nth order determinant has n^2 cofactors.

EXAMPLE 3.12 Find the minors and cofactors of

$$\det\mathbf{A} = \begin{vmatrix} 2 & -3 \\ 1 & 4 \end{vmatrix}.$$

Solution Inspection shows that $M_{11} = 4$, $M_{12} = 1$, $M_{21} = -3$, and $M_{22} = 2$. Using definition (12), the cofactors are seen to be

$$C_{11} = (-1)^{1+1} M_{11} = 4, \quad C_{12} = (-1)^{1+2} M_{12} = -1, \quad C_{21} = (-1)^{2+1} M_{21} = 3,$$

and $C_{22} = (-1)^{2+2} M_{22} = 2$. ■

Recognizing that the cofactors of the second order determinant

$$\det\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \text{are } C_{11} = a_{22},\ C_{12} = -a_{21},\ C_{21} = -a_{12},\ \text{and } C_{22} = a_{11},$$

we see from the definition $\det\mathbf{A} = a_{11}a_{22} - a_{21}a_{12}$ that $\det\mathbf{A}$ can be expressed in terms of these cofactors in four different ways:

expanding a second order determinant in terms of rows or columns

$\det\mathbf{A} = a_{11}C_{11} + a_{12}C_{12}$, using elements and cofactors from the first row of $\mathbf{A}$;

$\det\mathbf{A} = a_{21}C_{21} + a_{22}C_{22}$, using elements and cofactors from the second row of $\mathbf{A}$;

$\det\mathbf{A} = a_{11}C_{11} + a_{21}C_{21}$, using elements and cofactors from the first column of $\mathbf{A}$;

$\det\mathbf{A} = a_{12}C_{12} + a_{22}C_{22}$, using elements and cofactors from the second column of $\mathbf{A}$.

This has proved by direct calculation that the value of the general second order determinant $\det\mathbf{A}$ is given by the sum of the products of the elements and their associated cofactors in any row or column of the determinant. When the definition of a determinant is extended to the case $n > 2$ it will be seen that this same property remains true.

There are various ways of defining an nth order determinant, and from among these we have chosen to use one that involves a recursive process. More will be said about this recursive process, and how it can be used to evaluate the determinant, once the definition has been formulated.

Definition of a determinant of order *n*

The nth order determinant $\det\mathbf{A}$ in which the element a_{ij} has the associated cofactor C_{ij} for $i, j = 1, 2, \ldots, n$ is defined as

$$\det\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{j=1}^{n} a_{1j}C_{1j}. \tag{14}$$

Recalling the different ways in which a second order determinant can be evaluated, we see that the expansion of $\det\mathbf{A}$ in (14) is in terms of the elements and cofactors of the first row, so for conciseness this expansion is said to be in terms of the *elements of the first row*.

The recursive process enters this definition through the fact that each cofactor C_{1j} is a determinant of order $n-1$, as can be seen from (12), so each cofactor in turn can be expanded in terms of determinants of order $n-2$, and the process continued until determinants of order 2 are obtained that can then be calculated using (6).

EXAMPLE 3.13 Expand

$$\det\mathbf{A} = \begin{vmatrix} 1 & 4 & -1 \\ 2 & 0 & 3 \\ 1 & 2 & 1 \end{vmatrix}.$$

Solution To expand this third order determinant using (14), we must find the cofactors of the elements of the first row, so to do this we first find the minors and then use (13) to find the cofactors, as a result of which we find that

$$M_{11} = \begin{vmatrix} 0 & 3 \\ 2 & 1 \end{vmatrix} = -6, \quad \text{so } C_{11} = (-1)^{1+1}(-6) = -6$$

$$M_{12} = \begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = -1, \quad \text{so } C_{12} = (-1)^{1+2}(-1) = 1$$

$$M_{13} = \begin{vmatrix} 2 & 0 \\ 1 & 2 \end{vmatrix} = 4, \quad \text{so } C_{13} = (-1)^{1+3}(4) = 4.$$

As the elements of the first row are $a_{11} = 1$, $a_{12} = 4$, and $a_{13} = -1$, we find from (12) that

$$\det\mathbf{A} = (1)C_{11} + (4)C_{12} + (-1)C_{13} = (1)(-6) + (4)(1) + (-1)(4) = -6. \quad \blacksquare$$

The determinant associated with either an upper or a lower triangular matrix $\mathbf{A}$ of any order is easily expanded, because repeated application of (12) shows that it reduces to the product of the terms on the leading diagonal, so the expansion of the nth order upper triangular determinant with elements $a_{11}, a_{22}, \ldots, a_{nn}$ on its leading diagonal

$$\det\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ 0 & 0 & \cdots & \cdots \\ 0 & 0 & 0 & a_{nn} \end{vmatrix} = a_{11}a_{22}\ldots a_{nn}, \tag{15}$$

and a corresponding result is true for a lower triangular matrix.

Definition (14) can be used to prove that nth order determinants, like second order determinants, have the property that their value is given by the sum of the products of the elements and their cofactors in *any* row or column. This result, together with a generalization concerning the vanishing of the sum of the products of the elements in any row (or column) and the corresponding cofactors in a different row (or column), forms the next theorem. The details of the proof can be found in linear algebra texts, for example, [2.1], [2.5], [2.7], [2.9], but the method used has no other application in what is to follow, so the proof will be omitted.

THEOREM 3.3

Laplace expansion theorem

Laplace expansion theorem and an extension Let $\mathbf{A}$ be an $n \times n$ matrix with elements a_{ij}. Then,

(i) $\det\mathbf{A}$ can be expanded in terms of elements of its ith row and the cofactors C_{ij} of the ith row as

$$\det\mathbf{A} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}C_{ij}$$

for any fixed i with $1 \le i \le n$.

(ii) $\det\mathbf{A}$ can be expanded in terms of elements of its jth column and the cofactors C_{ij} of the jth column as

$$\det\mathbf{A} = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} = \sum_{j=1}^{n} a_{ij}C_{ij}$$

for any fixed j with $1 \le j \le n$.

(iii) The sum of the products of the elements of the ith row with the corresponding cofactors of the kth row is zero when $i \neq k$.

(iv) The sum of the products of the elements in the jth column with the corresponding cofactors of the kth column is zero when $j \neq k$. ■

Results (i) and (ii) are often used to advantage when a row or column contains many zeros, because if the determinant is expanded in terms of the elements of that row or column, the cofactors associated with each zero element need not be calculated.

Results (iii) and (iv) simply say that the sum of the products of the elements in any row (or column) with the corresponding cofactors in a *different* row (or column) is zero.

PIERRE SIMON LAPLACE (1749–1827)
A French mathematician of remarkable ability who made contributions to analysis, differential equations, probability, and celestial mechanics. He used mathematics as a tool with which to investigate physical phenomena, and made fundamental contributions to hydrodynamics, the propagation of sound, surface tension in liquids, and many other topics. His many contributions had a wide-ranging effect on the development of mathematics.

EXAMPLE 3.14

Verify Theorem 3.3(i) by expanding the determinant in Example 3.13 in terms of the elements of its second row. Use the determinant to check the result of Theorem 3.3(iii).

Solution The second row contains a zero element in its mid position, so the cofactor C_{22} associated with the zero element need not be calculated. The necessary cofactors in the second row that follow from the minors are

$$M_{21} = \begin{vmatrix} 4 & -1 \\ 2 & 1 \end{vmatrix} = 6 \quad \text{so } C_{21} = (-1)^{2+1}(6) = -6$$

$$M_{23} = \begin{vmatrix} 1 & 4 \\ 1 & 2 \end{vmatrix} = -2 \quad \text{so } C_{23} = (-1)^{2+3}(-2) = 2.$$

As $a_{21} = 2$ and $a_{23} = 3$, it follows from Theorem 3.3(i) that when $\det\mathbf{A}$ is expanded in terms of elements of its second row,

$$\det\mathbf{A} = (2)(-6) + (3)(2) = -6,$$

confirming the result obtained in Example 3.13.

As a particular case of Theorem 3.3(iii), let us show that the sum of the products of the cofactors in the first row of $\det\mathbf{A}$ and the corresponding elements in the third row is zero.

In Example 3.13 it was found that $C_{11} = -6$, $C_{12} = 1$, and $C_{13} = 4$, so as the elements of the third row are $a_{31} = 1$, $a_{32} = 2$, and $a_{33} = 1$, we have

$$a_{31}C_{11} + a_{32}C_{12} + a_{33}C_{13} = (-6)(1) + (2)(1) + (1)(4) = 0,$$

confirming the result of Theorem 3.3(iii) when the elements of row 3 and the cofactors of row 1 are used. ■

Determinants have a number of special properties that can be used to simplify their expansion, though their main uses are found elsewhere in mathematics, where determinants often characterize some important theoretical feature of a problem. The most important and useful of these properties are contained in the next theorem.

THEOREM 3.4

basic properties of determinants

Properties of determinants A determinant $\det\mathbf{A}$ has the following properties:

(i) If any row or column of a determinant $\det\mathbf{A}$ only contains zero elements, then $\det\mathbf{A} = 0$.

(ii) If $\mathbf{A}$ is a square matrix with the transpose $\mathbf{A}^{\mathrm{T}}$, then $\det\mathbf{A} = \det\mathbf{A}^{\mathrm{T}}$.

(iii) If each element of a row or column of a square matrix $\mathbf{A}$ is multiplied by a constant k, then the value of the determinant is $k\det\mathbf{A}$.

(iv) If two rows (or columns) of a square matrix are interchanged, the sign of the determinant is changed.

(v) If any two rows or columns of a square matrix $\mathbf{A}$ are proportional, then $\det\mathbf{A} = 0$.

(vi) Let the square matrix $\mathbf{A}$ be such that each element a_{ij} of the ith row (or the jth column) can be written as $a_{ij} = a_{ij}^{(1)} + a_{ij}^{(2)}$. Then if $\mathbf{A}_1$ is the matrix derived from $\mathbf{A}$ by replacing its ith row (or jth column) by the elements $a_{ij}^{(1)}$ and $\mathbf{A}_2$ is the matrix derived from $\mathbf{A}$ by replacing its ith row (or jth column) by the elements $a_{ij}^{(2)}$,

$$\det\mathbf{A} = \det\mathbf{A}_1 + \det\mathbf{A}_2.$$

(vii) The addition of a multiple of a row (or column) of a determinant to another row (or column) of the determinant leaves the value of the determinant unchanged.

(viii) Let $\mathbf{A}$ and $\mathbf{B}$ be two $n \times n$ matrix, then

$$\det(\mathbf{AB}) = \det\mathbf{A}\,\det\mathbf{B}.$$

Proof

(i) The result follows by expanding the determinant in terms of the row or column that only contains zero elements.

(ii) The result follows from the fact that expanding $\det\mathbf{A}$ in terms of the elements of its first row is the same as expanding $\det\mathbf{A}^{\mathrm{T}}$ in terms of the elements of its first column.

(iii) The result follows by expanding the determinant in terms of the row or column in which each element has been multiplied by the constant k, because k appears as a factor in each term, so the result becomes $k\det\mathbf{A}$.

(iv) The proof is by induction, starting with a second order determinant for which the result can be seen to be true from definition (6). To proceed with an inductive proof we assume the results to be true for a determinant of order $r-1$, and show it must be true for a determinant of order r. Expand a row of a determinant of order r in terms of the elements of a row (or column) that has not been interchanged. Then, by hypothesis, as the cofactors are determinants of order $r-1$, their signs will all be reversed. This establishes that if the hypothesis is true for a determinant of order $r-1$ it must also be true for a determinant of order r. As the result is true for $r=2$, it follows by induction that it is true for all integers $r>2$, and the result is proved.

(v) If the value of the determinant is $\det\mathbf{A}$, and one row is k times another, then from (ii) by removing the factor k from the row the value of the determinant will be $k\det\mathbf{A}_1$, where $\mathbf{A}_1$ is now a determinant with two identical rows. From (ii), interchanging two rows changes the sign of the determinant, but the rows are identical, leaving the determinant invariant, so $\det\mathbf{A}_1=0$. A similar argument shows the result to be true when two columns are proportional, so the result is proved.

(vi) The result is proved directly by expanding the determinant in terms of the elements of the ith row (or the jth column).

(vii) Let the square matrix $\mathbf{B}$ be obtained from $\mathbf{A}$ by adding k times the ith row (or a column) to the jth row (or column). Then from (iii) and (vi),

$$\det\mathbf{B}=\det\mathbf{A}+k\det\mathbf{C},$$

where $\mathbf{C}$ is obtained from $\mathbf{A}$ by replacing the ith row (or column) by the jth row (or column). As $\det\mathbf{C}$ has two identical rows (or columns), it follows from (v) that $\det\mathbf{C}=0$, so $\det\mathbf{B}=\det\mathbf{A}$ and the result is proved.

(viii) A proof of this result will be given later after the introduction of elementary row operation matrices. ∎

Cramer's rule, which was first encountered when seeking the solution of the two equations in (7), can be extended to a system of n equations in a very straightforward manner, and it takes the following form.

Cramer's rule

Cramer's rule for a system of *n* equations in *n* unknowns

The solution of the system of n equations in the n unknowns $x_1, x_2, \ldots, x_n$

$$\begin{aligned}
a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n&=b_1\\
a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n&=b_2\\
\cdots\cdots&\\
a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n&=b_n
\end{aligned}$$

is given by

$$x_i = \det \mathbf{A}_i / \det \mathbf{A} \quad \text{for } i = 1, 2, \ldots, n,$$

where $\det \mathbf{A}$ is the determinant of the coefficient matrix with elements a_{ij}, and $\det \mathbf{A}_i$ is the determinant obtained from the coefficient matrix by replacing its ith column by the column containing the number $b_1, b_2, \ldots, b_n$.

The justification for Cramer's rule in this more general form will be postponed until after the introduction of inverse matrices, when a simple proof can be given. Cramer's rule is mainly of theoretical importance and, in general, it should not be used to solve equations when $n > 3$. This is because the number of multiplications required to evaluate a determinant of order n is $(n-1)n!$, so to solve for n unknowns $(n+1)$ determinants must be evaluated leading to a total of $(n^2-1)n!$ multiplications, and this calculation becomes excessive when $n > 3$. An efficient way of solving large systems by means of elimination is given in Chapter 19.

EXAMPLE 3.15

Use Cramer's rule to solve

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 1 \\ 2x_1 + x_2 - 2x_3 &= 3 \\ -x_1 + 3x_2 + 4x_3 &= -2. \end{aligned}$$

Solution The determinants involved are

$$\det \mathbf{A} = \begin{vmatrix} 1 & -2 & 1 \\ 2 & 1 & -2 \\ -1 & 3 & 4 \end{vmatrix} = 29, \quad \det \mathbf{A}_1 = \begin{vmatrix} 1 & -2 & 1 \\ 3 & 1 & -2 \\ -2 & 3 & 4 \end{vmatrix} = 37$$

$$\det \mathbf{A}_2 = \begin{vmatrix} 1 & 1 & 1 \\ 2 & 3 & -2 \\ -1 & -2 & 4 \end{vmatrix} = 1, \quad \det \mathbf{A}_3 = \begin{vmatrix} 1 & -2 & 1 \\ 2 & 1 & 3 \\ -1 & 3 & -2 \end{vmatrix} = -6,$$

so $x_1 = 37/29$, $x_2 = 1/29$, and $x_3 = -6/29$. ■

A purely algebraic approach to the study of determinants and their properties is to be found in reference [2.8], and many examples of their applications are given in references [2.11] and [2.12].

Summary

This section has extended to an nth order determinant the basic notion of a second order determinant that was reviewed in Chapter 1, and then established its most important properties. The Laplace expansion formulas that were established are of theoretical importance, but it will be seen later that the practical evaluation of a determinant is most easily performed by reducing the $n \times n$ matrix associated with a determinant to its echelon form.

EXERCISES 3.3

In Exercises 1 through 4 find $\det\mathbf{A}$.

1. $\det\mathbf{A} = \begin{vmatrix} 2 & 1 & -1 \\ 0 & 4 & 3 \\ 3 & 2 & -2 \end{vmatrix}$.

2. $\det\mathbf{A} = \begin{vmatrix} -1 & 2 & 1 \\ 1 & 3 & 2 \\ -4 & 1 & 2 \end{vmatrix}$.

3. $\det\mathbf{A} = \begin{vmatrix} 2 & 4 & -3 \\ -2 & 1 & 0 \\ 5 & -2 & 4 \end{vmatrix}$.

4. $\det\mathbf{A} = \begin{vmatrix} 4 & 0 & 0 \\ -2 & \cos x & -\sin x \\ 5 & \sin x & \cos x \end{vmatrix}$.

5. Given that

$$\det\mathbf{A} = \begin{vmatrix} -3 & 1 & 4 \\ 2 & -1 & 5 \\ 4 & 2 & 5 \end{vmatrix} = 87,$$

confirm by direct calculation that (a) interchanging the first and last rows changes the sign of $\det\mathbf{A}$ and (b) interchanging the second and third columns changes the sign of $\det\mathbf{A}$.

6. Given that

$$\det\mathbf{A} = \begin{vmatrix} 2 & 1 & 3 \\ 5 & -2 & 2 \\ -1 & 1 & 3 \end{vmatrix} = -24,$$

confirm by direct calculation that (a) adding twice row two to row three leaves $\det\mathbf{A}$ unchanged and (b) subtracting three times column three from column one leaves $\det\mathbf{A}$ unchanged.

Establish the results in Exercises 7 through 12 without a direct expansion of the determinant by using the properties listed in Theorem 3.4.

7. $\begin{vmatrix} 1+a & a & a \\ b & 1+b & b \\ c & c & 1+c \end{vmatrix} = (1+a+b+c)$.

8. $\begin{vmatrix} 1 & a & b+c \\ 1 & b & c+a \\ 1 & c & a+b \end{vmatrix} = 0$.

9. $\begin{vmatrix} a^2 & b^2 & c^2 \\ a & b & c \\ 1 & 1 & 1 \end{vmatrix} = (a-b)(a-c)(b-c)$.

10. $\begin{vmatrix} x^2+a^2 & ab & ac \\ ab & x^2+b^2 & bc \\ ac & cb & x^2+c^2 \end{vmatrix} = x^4(x^2+a^2+b^2+c^2)$.

11. $\begin{vmatrix} 1 & a & b \\ a & 1 & b \\ a & b & 1 \end{vmatrix} = (a+b+1)(a-1)(b-1)$.

12. $\begin{vmatrix} k & 1 & 1 & 1 \\ 1 & k & 1 & 1 \\ 1 & 1 & k & 1 \\ 1 & 1 & 1 & k \end{vmatrix} = (k+3)(k-1)^3$.

In Exercises 13 and 14 use Cramer's rule to solve the system of equations.

13. $\begin{aligned} 2x_1 - 3x_2 + x_3 &= 4 \\ x_1 + 2x_2 - 2x_3 &= 1 \\ 3x_1 + x_2 - 2x_3 &= -2. \end{aligned}$

14. $\begin{aligned} 3x_1 + x_2 + 2x_3 &= 5 \\ 2x_1 - 4x_2 + 3x_3 &= -3 \\ x_1 + 2x_2 + 4x_3 &= 2. \end{aligned}$

15. Let $P(\lambda)$ be given by

$$P(\lambda) = \begin{vmatrix} 3-\lambda & 0 & 1 \\ 2 & 2-\lambda & 2 \\ 4 & 2 & 1-\lambda \end{vmatrix},$$

where λ is a parameter. Expand the determinant to find the form of the polynomial $P(\lambda)$ and use the result to find for what values of λ the determinant vanishes.

16. Let $P(\lambda)$ be given by

$$P(\lambda) = \begin{vmatrix} 4-\lambda & 0 & 1 \\ 1 & -\lambda & 1 \\ -1 & -2 & 2-\lambda \end{vmatrix},$$

where λ is a parameter. Expand the determinant to find the form of the polynomial $P(\lambda)$ and use the result to find for what values of λ the determinant vanishes.

17. Given that

$$\mathbf{A} = \begin{bmatrix} -3 & 0 & 4 \\ 1 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix},$$

calculate $\det(\mathbf{AB})$, $\det\mathbf{A}$, $\det\mathbf{B}$, and hence verify that $\det(\mathbf{AB}) = \det\mathbf{A}\det\mathbf{B}$.

3.4 Elementary Row Operations, Elementary Matrices, and Their Connection with Matrix Multiplication

To motivate what is to follow we will examine the processes involved when solving by elimination the system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots& \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{16}$$

though later more will need to be said about the details of this important problem, and how it is influenced by the number of equations m and the number of unknowns n.

Elementary Row Operations

the three basic types of elementary row operation

The three types of elementary row operations used when solving equations (16) by elimination are:

TYPE I The interchange of two equations

TYPE II The scaling of an equation by a nonzero constant

TYPE III The addition of a scalar multiple of an equation to another equation

In matrix notation the system of equations (16) becomes

$$\mathbf{Ax} = \mathbf{b}, \tag{17}$$

where $\mathbf{A} = [a_{ij}]$ is an $m \times n$ matrix, $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$, and $\mathbf{b} = [b_1, b_2, \ldots, b_m]^{\mathrm{T}}$. The three elementary row operations of types I to III that can be performed on the *equations* in (16) can be interpreted as the corresponding operations performed on the *rows* of the matrices $\mathbf{A}$ and $\mathbf{b}$. This is equivalent to performing these same operations on the rows of the new matrix denoted by $(\mathbf{A}, \mathbf{b})$, defined as

$$(\mathbf{A}, \mathbf{b}) = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ & & \cdots\cdots & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right], \tag{18}$$

that has m rows and $n + 1$ columns and is obtained by inserting the column vector $\mathbf{b}$ containing the nonhomogeneous terms on the right of matrix $\mathbf{A}$.

the augmented matrix

When considering the system of linear equations in (16), matrix $(\mathbf{A}, \mathbf{b})$ is called the **augmented matrix** associated with the system. The separation of the last column in (18) by a vertical dashed line is to indicate *partitioning* of the matrix to show that the elements of the last column are not elements of the coefficient matrix $\mathbf{A}$.

We are now in a position to introduce a notation for the three *elementary row operations* that are necessary when using an elimination process to find the solution of a system of equations in matrix form (ordinary or augmented).

Elementary row operations

The three **elementary row operations** that may be performed on a matrix are:

(i) The interchange of the ith and jth rows, which will be denoted by $R\{i \to j, j \to i\}$.

(ii) The replacement of each element in the ith row by its product with a nonzero constant α, which will be denoted by $R\{(\alpha)i \to i\}$.

(iii) The replacement of each element of the jth row by the sum of β times the corresponding element in the ith row and the element in the jth row, which will be denoted by $R\{(\beta)i + j \to j\}$.

EXAMPLE 3.16 To illustrate the elementary row operations, we consider the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 6 & 4 & -3 & 2 \\ 2 & 0 & 1 & 7 & 4 \\ 5 & 2 & 8 & 2 & 3 \end{bmatrix}.$$

An example of an elementary row operation of type (i) performed on $\mathbf{A}$ is provided by $R\{1 \to 3, 3 \to 1\}$. This requires rows 1 and 3 to be interchanged to give the new matrix

$$R\{1 \to 3, 3 \to 1\}\mathbf{A} = \begin{bmatrix} 5 & 2 & 8 & 2 & 3 \\ 2 & 0 & 1 & 7 & 4 \\ 1 & 6 & 4 & -3 & 2 \end{bmatrix}.$$

An example of an elementary row operation of type (ii) performed on $\mathbf{A}$ is provided by $R\{(-3)1 \to 1\}$. This requires each element in row 1 to be multiplied by -3 to give the new matrix

$$R\{(-3)1 \to 1\}\mathbf{A} = \begin{bmatrix} -3 & -18 & -12 & 9 & -6 \\ 2 & 0 & 1 & 7 & 4 \\ 5 & 2 & 8 & 2 & 3 \end{bmatrix}.$$

An example of an elementary row operation of type (iii) performed on $\mathbf{A}$ is provided by $R\{(4)1 + 2 \to 2\}$, which requires the elements of row 1 to be multiplied by 4 and then added to the corresponding elements of row 2 to give the new matrix

$$R\{(4)1 + 2 \to 2\}\mathbf{A} = \begin{bmatrix} 1 & 6 & 4 & -3 & 2 \\ 6 & 24 & 17 & -5 & 12 \\ 5 & 2 & 8 & 2 & 3 \end{bmatrix}.$$

■

A sequence of elementary row operations performed on the augmented matrix $(\mathbf{A}, \mathbf{b})$ will lead to a different augmented matrix $(\mathbf{A}', \mathbf{b}')$. However, as this is equivalent to performing the corresponding sequence of operations on the actual equations in (16), although $(\mathbf{A}, \mathbf{b})$ and $(\mathbf{A}', \mathbf{b}')$ will look different, the interpretation of $(\mathbf{A}', \mathbf{b}')$ in terms of the solution of the system of equations in (16) will, of course, be the same as that of $(\mathbf{A}, \mathbf{b})$. It will be seen later that the purpose of carrying out these operations on a matrix is to simplify it while leaving its essential

algebraic structure unaltered, e.g., without changing the solution $x_1, \ldots, x_n$ of the corresponding system of equations.

The definition that now follows is a consequence of the equivalence, in terms of equations (16), of matrix $(\mathbf{A}, \mathbf{b})$ and any matrix $(\mathbf{A}', \mathbf{b}')$ that can be derived from it by means of a sequence of elementary row operations, though the definition applies to matrices in general, and not only to augmented matrices.

Row equivalence of matrices

Two $m \times n$ matrices will be said to be **row equivalent** if one can be obtained from the other by means of a sequence of elementary row operations. Row equivalence between matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted by writing $\mathbf{A} \sim \mathbf{B}$.

The row equivalence of matrices has the useful properties listed in the following theorem.

THEOREM 3.5 **Reflexive, symmetric, and transitive properties of row equivalence**

(i) Every $m \times n$ matrix $\mathbf{A}$ is row equivalent to itself (*reflexive* property).

(ii) Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ matrices. Then if $\mathbf{A}$ is row equivalent to $\mathbf{B}$, $\mathbf{B}$ is row equivalent to $\mathbf{A}$ (*symmetric* property).

(iii) Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be $m \times n$ matrices. Then if matrix $\mathbf{A}$ is row equivalent to $\mathbf{B}$ and $\mathbf{B}$ is row equivalent to $\mathbf{C}$, $\mathbf{A}$ is row equivalent to $\mathbf{C}$ (*transitive* property).

Proof

(i) The property is self-evident.

(ii) To establish this property we must show the three elementary row operations involved are reversible. In the case of elementary row operations of type (i) the result follows from the fact that if an application of the operation $R\{i \to j, j \to i\}$ to matrix $\mathbf{A}$ yields a new matrix $\mathbf{B}$, an application of the operation $R\{j \to i, i \to j\}$ to matrix $\mathbf{B}$ generates the original matrix $\mathbf{A}$.

Similarly, in the case of elementary row operations of type (ii), if an application of the operation $R\{(\alpha)i \to i\}$ to matrix $\mathbf{A}$ yields a new matrix $\mathbf{B}$, an application of the operation $R\{(1/\alpha)i \to i\}$ to matrix $\mathbf{B}$ reproduces the original matrix $\mathbf{A}$.

Finally we consider the case of elementary row operations of type (iii). If an application of the operation $R\{(\beta)i + j \to j\}$ to matrix $\mathbf{A}$ yields a new matrix $\mathbf{B}$, an application of the operation $R\{(-\beta)i + j \to j\}$ to $\mathbf{B}$ returns the original matrix $\mathbf{A}$. Taken together these results establish property (ii).

(iii) Using property (ii) in (iii) establishes the row equivalence first of $\mathbf{A}$ and $\mathbf{B}$, and then of $\mathbf{B}$ and $\mathbf{C}$, and hence of $\mathbf{A}$ and $\mathbf{C}$, so property (iii) is proved. ■

Let us now define what are called *elementary matrices* and examine the effect they have when used to premultiply a matrix.

Elementary matrices

An $n \times n$ **elementary matrix** is any matrix that is obtained from an $n \times n$ unit matrix $\mathbf{I}$ by performing a single elementary row operation.

The following concise notation will be used to identify the elementary matrices that correspond to each of the three elementary row operations.

the three basic types of elementary matrix

TYPE I $\mathbf{E}_{ij}$ will denote the elementary matrix obtained from the unit matrix $\mathbf{I}$ by interchanging its ith and jth rows.

TYPE II $\mathbf{E}_i(c)$ will denote the matrix obtained from the unit matrix $\mathbf{I}$ by multiplying its ith row by the nonzero scalar c.

TYPE III $\mathbf{E}_{ij}(c)$ will denote the matrix obtained from the unit matrix $\mathbf{I}$ by adding c times its ith row to its jth row.

EXAMPLE 3.17 Let $\mathbf{I}$ be the 3×3 unit matrix. Then

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{E}_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{E}_3(4) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \quad \text{and}$$

$$\mathbf{E}_{13}(5) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 5 & 0 & 1 \end{bmatrix}.$$

Determinants of Elementary Matrices

It follows directly from the definitions of elementary matrices that:

(a) The determinant of an elementary matrix of Type I is -1, because two rows of a unit matrix have been interchanged so, in terms of $\mathbf{E}_{ij}$, we have $\det(\mathbf{E}_{ij}) = -1$.

(b) The determinant of an elementary matrix of Type II in which a row is multiplied by a nonzero constant c is c, because a row of a unit matrix has been multiplied by c so, in terms of $\mathbf{E}_i(c)$, we have $\det(\mathbf{E}_i(c)) = c$.

(c) The determinant of an elementary matrix of Type III in which c times one row has been added to another row is 1, because the addition of a multiple of a row of a unit matrix to another row leaves its value unchanged so, in terms of $\mathbf{E}_{ij}(c)$, we have $\det(\mathbf{E}_{ij}(c)) = 1$.

The next theorem shows that premultiplication of a matrix $\mathbf{A}$ by an elementary matrix $\mathbf{E}$ that is conformable for multiplication performs on $\mathbf{A}$ the same elementary row operation that was used to generate $\mathbf{E}$ from $\mathbf{I}$.

THEOREM 3.6 **Row operations performed by elementary matrices** Let $\mathbf{E}$ be an $m \times m$ elementary matrix produced by performing an elementary row operation on the unit matrix $\mathbf{I}$, and let $\mathbf{A}$ be an $m \times n$ matrix. Then the matrix product $\mathbf{EA}$ is the matrix that is obtained when the row operation that generated $\mathbf{E}$ from $\mathbf{I}$ is performed on $\mathbf{A}$.

Proof The proof of the theorem follows directly from the definition of a matrix product and the fact that, with the exception of the ith element in the ith row of $\mathbf{I}$, which is 1, all the other elements in that row are zero. So if $\mathbf{E}$ is the elementary matrix obtained from $\mathbf{I}$ by replacing the element 1 in its ith row by α, the result of the matrix product $\mathbf{EA}$ will be that the elements in the ith row of $\mathbf{A}$ will be multiplied by α. As the form of argument used to establish the effect on $\mathbf{A}$ of premultiplication by $\mathbf{P}$ to form $\mathbf{PA}$ can also be employed when the other two elementary row operations are used to generate an elementary matrix $\mathbf{E}$, the details will be left as an exercise. ■

EXAMPLE 3.18 Let $\mathbf{A}$ be the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 5 \\ 1 & 3 & 7 \\ 6 & 1 & 2 \end{bmatrix}.$$

If we use the notation for elementary matrices, and introduce the elementary matrix $\mathbf{E}_{23}$ from Example 3.17 obtained by interchanging the last two rows of $\mathbf{I}_3$, a routine calculation shows that

$$\mathbf{E}_{23}\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 1 & 3 & 7 \\ 6 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 5 \\ 6 & 1 & 2 \\ 1 & 3 & 7 \end{bmatrix},$$

so the product $\mathbf{E}_{23}\mathbf{A}$ has indeed interchanged the last two rows of $\mathbf{A}$.

Similarly, again using the elementary matrices in Example 3.17, it is easily checked that $\mathbf{E}_3(4)\mathbf{A}$ multiplies the elements in the third row of $\mathbf{A}$ by 4, while $\mathbf{E}_{13}(5)\mathbf{A}$ adds five times the first row of $\mathbf{A}$ to the last row. ■

The main use of Theorem 3.6 is to be found in the theory of matrix algebra, and in the justification it provides for various practical methods that are used when working with matrices. This is because when solving purely numerical problems the necessary row operations need only be performed on the rows of the augmented matrix instead of on the system of equations itself.

Typical uses of the theorem will occur later after a discussion of the linear independence of equations, the definition of what is called the *rank* of a matrix, and the introduction of the inverse of an $n \times n$ matrix $\mathbf{A}$. In this last case, the results of the theorem will be used to provide an elementary method by which what is called the inverse matrix of an $n \times n$ matrix can be obtained when n is small.

Summary This section introduced the three types of elementary row operations that are used when manipulating matrices together with the corresponding three types of elementary matrix that can be used to perform elementary row operations.

3.5 The Echelon and Row-Reduced Echelon Forms of a Matrix

We now use the row equivalence of matrices to reduce a matrix $\mathbf{A}$ to one of two slightly different but related standard forms called, respectively, its *echelon form* and its *row-reduced echelon form*. It is helpful to introduce these two new concepts by considering the solution of the system of m equations in n unknowns introduced in (16) and written in an equivalent but more condensed form as $(\mathbf{A}, \mathbf{b})$, where

$$(\mathbf{A}, \mathbf{b}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}, \qquad (19)$$

because this is equivalent to the full matrix equation $\mathbf{Ax} = \mathbf{b}$.

Echelon and row-reduced echelon forms of a matrix

echelon and row-reduced echelon forms

A matrix **A** is said to be in **echelon form** if:

(i) The first nonzero element in each row, called its *leading entry*, is 1;

(ii) In any two successive rows i and $i + 1$ that do not consist entirely of zeros the leading element in the $(i + 1)$th row lies to the right of the leading element in ith row;

(iii) Any rows that consist entirely of zeros lie at the bottom of the matrix.

Matrix **A** is said to be in **row-reduced echelon form** if, in addition to conditions (i) to (iii), it is also true that

(iv) In a column that contains the leading entry of a row, all the other elements are zero.

In summary, this definition means that a matrix **A** is in *echelon form* if the first nonzero entry in any row is a 1, the entry appears to the right of the first nonzero entry in the row above, and all rows of zeros lie at the bottom of the matrix. Furthermore, matrix **A** is in *row-reduced echelon form* if, in addition to these conditions, the first nonzero entry in any row is the only nonzero entry in the column containing that entry.

EXAMPLE 3.19

The following matrices are in *echelon form*:

$$\begin{bmatrix} 1 & 0 & 5 & 7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 5 & 2 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The matrices

$$\begin{bmatrix} 0 & 1 & 0 & 2 & 0 & 5 & 0 \\ 0 & 0 & 1 & 1 & 0 & 3 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 9 & 2 \\ 0 & 1 & 0 & 2 & 3 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

are in *row-reduced echelon form*. ■

Rules for the reduction of a matrix to echelon form

rules for finding the echelon form

The reduction of the $m \times n$ matrix to its echelon form is accomplished by means of the following steps:

1. Find the row whose first nonzero element is furthest to the left and, if necessary, move it into row 1; if there is more than one such row, choose the row whose first nonzero element has the largest absolute value.
2. Scale row 1 to make its leading entry 1.
3. Subtract multiples of row 1 from the $m - 1$ rows below it to reduce to zero all entries that lie below the leading entry in the first column.
4. In the $m - 1$ rows below row 1, find the row whose first nonzero entry is furthest to the left and, if necessary, move it into row 2; if there is more

than one such row, choose the row whose first nonzero entry has the largest absolute value.

5. Scale row 2 to make its leading entry 1.
6. Subtract multiples of row 2 from the $m - 2$ rows below it to reduce to zero all entries in the column below the leading entry in row 2.
7. Continue this process until either the first nonzero entry in the mth row is 1, or a stage is reached at which all subsequent rows consist entirely of zeros.
8. The matrix is then in its echelon form.

Remark

The selection in Step 1, and the steps corresponding to Step 4, of a row whose first nonzero entry has the largest magnitude is made to reduce computational errors, and is not necessary mathematically. This criterion is introduced to ensure that the elimination procedure does not use an unnecessary scaling of a nonzero entry of small absolute magnitude to reduce to zero an entry of large absolute magnitude.

rules for finding the row-reduced echelon form

Rules for the reduction of a matrix to row-reduced echelon form

1. Proceed as in the reduction of a matrix to echelon form, but when steps equivalent to Step 6 are reached, in addition to subtracting multiples of the row containing a leading entry 1 from the rows below to reduce to zero all elements in the column below the leading entry, this same process must be repeated to reduce to zero all elements in the column above the leading entry.
2. An equivalent approach is first to reduce the matrix to echelon form and then, starting with row 2 and working downwards, to subtract multiples of successive rows from the rows above to generate columns with leading entries to ones with the single nonzero entry 1.

Each of these methods reduces a matrix to its row-reduced echelon form.

The row equivalence of a matrix with either its echelon or its row-reduced echelon form means that the different-looking systems of equations represented by these three matrices all have identical solution sets. The simplified structure of the row echelon and row-reduced echelon forms of the original augmented matrix makes the solution of the associated system of equations particularly easy, as can be seen from the following examples.

EXAMPLE 3.20 Reduce the following matrix to its echelon and its row-reduced echelon form:

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 2 & 4 & 8 & 2 & 4 \\ 1 & 2 & 4 & 2 & 2 \\ 1 & 3 & 6 & 1 & 5 \end{bmatrix}.$$

Solution

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 2 & 4 & 8 & 2 & 4 \\ 1 & 2 & 4 & 2 & 2 \\ 1 & 3 & 6 & 1 & 5 \end{bmatrix} \underset{\substack{\text{switch rows}\\ \text{2 and 1}}}{\sim} \begin{bmatrix} 2 & 4 & 8 & 2 & 4 \\ 0 & 1 & 2 & 0 & 3 \\ 1 & 2 & 4 & 2 & 2 \\ 1 & 3 & 6 & 1 & 5 \end{bmatrix}$$

$$\underset{\text{divide row 1 by 2}}{\sim} \begin{bmatrix} 1 & 2 & 4 & 1 & 2 \\ 0 & 1 & 2 & 0 & 3 \\ 1 & 2 & 4 & 2 & 2 \\ 1 & 3 & 6 & 1 & 5 \end{bmatrix} \underset{\substack{\text{subtract row 1}\\ \text{from rows 3 and 4}}}{\sim} \begin{bmatrix} 1 & 2 & 4 & 1 & 2 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 3 \end{bmatrix}$$

$$\underset{\substack{\text{subtract row 2}\\ \text{from row 4}}}{\sim} \begin{bmatrix} 1 & 2 & 4 & 1 & 2 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and the matrix is now in echelon form.

Having already obtained the echelon form of the matrix, we now use it to obtain the row-reduced echelon form. We already have

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 2 & 4 & 8 & 2 & 4 \\ 1 & 2 & 4 & 2 & 2 \\ 1 & 3 & 6 & 1 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 4 & 1 & 2 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \underset{\substack{\text{subtract twice row 2}\\ \text{from row 1}}}{\sim}$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & -4 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \underset{\substack{\text{subtract row 3}\\ \text{from row 1}}}{\sim} \begin{bmatrix} 1 & 0 & 0 & 0 & -4 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and the matrix is now in its row-reduced echelon form. ■

EXAMPLE 3.21 Solve the system of equations

$$\begin{aligned} x_2 + 2x_3 &= 3 \\ 2x_1 + 4x_2 + 8x_3 + 2x_4 &= 4 \\ x_1 + 2x_2 + 4x_3 + 2x_4 &= 2 \\ x_1 + 3x_2 + 6x_3 + x_4 &= 5. \end{aligned}$$

Solution The augmented matrix $(\mathbf{A}, \mathbf{b})$ for this system is the matrix in Example 3.20 that was shown to be equivalent to the row-reduced echelon form

$$\left[\begin{array}{cccc:c} 1 & 0 & 0 & 0 & -4 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

If we recall that the first four columns of this matrix contain the coefficients of x_1, x_2, x_3, and x_4, while the last column contains the nonhomogeneous terms, the matrix implies the much simpler system of equations

$$x_4 = 0, \quad x_2 + 2x_3 = 3, \quad \text{and} \quad x_1 = -4.$$

As there are only three equations connecting four unknowns, it follows that in the second equation either x_2 or x_3 can be assigned arbitrarily, so if we choose to set $x_3 = k$ (an arbitrary number), the solution set of the system in terms of the parameter k becomes

$$x_1 = -4, \quad x_2 = 3 - 2k, \quad x_3 = k, \quad \text{and} \quad x_4 = 0.$$

The same solution could have been obtained from the echelon form of the matrix

$$\left[\begin{array}{cccc:c} 1 & 2 & 4 & 1 & 2 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

because this implies the system of equations

$$x_1 + 2x_2 + 4x_3 + x_4 = 2, \quad x_2 + 2x_3 = 3, \quad \text{and} \quad x_4 = 0.$$

Starting from the last equation we find $x_4 = 0$, and setting $x_3 = k$ in the middle equation gives, as before, $x_2 = 3 - 2k$. Finally, substituting x_2, x_3, and x_4 in the first equation gives $x_1 = -4$. This process of arriving at a solution of a system of equations whose coefficient matrix is in upper triangular form is called **back substitution**.

back substitution

It should be noticed that the system of equations would have had *no solution* if the row-reduced echelon form had been

$$\left[\begin{array}{cccc:c} 1 & 0 & 0 & 0 & -4 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{array}\right].$$

This is because although the equations corresponding to the first three rows of this matrix would have been the same as before, the fourth row implies $0 = 5$, which is impossible. This corresponds to a system of equations where one equation contradicts the others, so that no solution is possible. ■

Summary

This section defined two related types of fundamental matrix that can be obtained from a general matrix by means of elementary row operations. The first was a reduction to echelon form and the second, derived from the first form, was a reduction to row-reduced echelon form. Each of the reduced forms retains the essential properties of the original matrix, while simplifying the task of solving the associated system of linear algebraic equations.

EXERCISES 3.5

Let **P**, **Q**, and **R** be the matrices

$$\mathbf{P} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In Exercises 1 through 4 verify by direct calculation that (a) premultiplication by **P** multiplies row 1 by 3; (b) premultiplication by **Q** interchanges rows 1 and 3; and (c) premultiplication by **R** adds twice row 2 to row 1.

1. $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 2 & 4 \end{bmatrix}$.

2. $\begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 3 \\ 3 & 0 & 7 \end{bmatrix}$.

3. $\begin{bmatrix} 4 & 0 & 1 \\ 2 & 0 & 3 \\ 1 & 2 & 5 \end{bmatrix}$.

4. $\begin{bmatrix} 9 & 1 & 3 \\ 2 & 4 & 7 \\ 1 & 2 & 2 \end{bmatrix}$.

In Exercises 5 and 6 write down the required elementary matrices.

5. When $\mathbf{I}$ is the 3×3 unit matrix, write down $\mathbf{E}_{12}$, $\mathbf{E}_2(3)$, and $\mathbf{E}_{12}(6)$.

6. When $\mathbf{I}$ is the 4×4 unit matrix, write down $\mathbf{E}_{41}$, $\mathbf{E}_4(3)$, and $\mathbf{E}_{23}(4)$.

In Exercises 7 through 12, reduce the given matrices to their row-reduced echelon form.

7. $\begin{bmatrix} 0 & 3 & 4 & 1 \\ 3 & 1 & 2 & 2 \\ 1 & 5 & 2 & 1 \end{bmatrix}$.

8. $\begin{bmatrix} 4 & 1 & 3 & 1 & 3 \\ 2 & 1 & 1 & 2 & 0 \\ 3 & 2 & 1 & 1 & 0 \end{bmatrix}$.

9. $\begin{bmatrix} 4 & -2 & 2 & 3 & 1 \\ 2 & 0 & 0 & 3 & 2 \\ 4 & 1 & 2 & 5 & 1 \end{bmatrix}$.

10. $\begin{bmatrix} 3 & 2 & 1 & 1 \\ 2 & 5 & 1 & 2 \\ 3 & 1 & 1 & 3 \\ 0 & 1 & 3 & 4 \\ 2 & 1 & 3 & 1 \end{bmatrix}$.

11. $\begin{bmatrix} 2 & 2 & 4 & 1 & 4 \\ 1 & 1 & 3 & 2 & 1 \\ 3 & 2 & 5 & 1 & 4 \\ 1 & 0 & 3 & 1 & 2 \end{bmatrix}$.

12. $\begin{bmatrix} 3 & 2 & 3 & 2 \\ 3 & 7 & 1 & -1 \\ 5 & 1 & 1 & 3 \end{bmatrix}$.

In Exercises 13 through 18, reduce the given augmented matrices to their row-reduced echelon form and, where appropriate, use the result to solve the related system of equations in terms of an appropriate number of the unknowns $x_1, x_2, \ldots$.

13. $\left[\begin{array}{ccc:c} 2 & 3 & 1 & 0 \\ 1 & 3 & 1 & 4 \\ 6 & 9 & 4 & 8 \end{array}\right]$.

14. $\left[\begin{array}{ccc:c} 2 & 1 & 1 & 0 \\ 2 & 3 & 1 & 4 \\ 4 & 9 & 4 & 8 \end{array}\right]$.

15. $\left[\begin{array}{cccc:c} 0 & 2 & 1 & 1 & 1 \\ 1 & 3 & 1 & 2 & 1 \\ 3 & 9 & 4 & 3 & 0 \end{array}\right]$.

16. $\left[\begin{array}{cccc:c} 2 & 1 & 0 & 2 & 1 \\ 1 & 3 & 1 & 4 & 2 \\ 2 & 1 & 2 & 3 & 1 \\ 4 & 7 & 4 & 11 & 7 \end{array}\right]$.

17. $\left[\begin{array}{ccccc:c} 1 & 0 & 1 & 0 & 2 & 0 \\ 2 & 2 & 6 & 0 & 6 & 0 \\ 1 & 0 & 1 & 1 & 6 & 0 \\ 3 & 2 & 7 & 0 & 8 & 2 \end{array}\right]$.

18. $\left[\begin{array}{cccc:c} 3 & 0 & 6 & 0 & 6 \\ 1 & 1 & 5 & 1 & 9 \\ 2 & 0 & 4 & 2 & 10 \end{array}\right]$.

3.6 Row and Column Spaces and Rank

The reduction of an $m \times n$ matrix $\mathbf{A}$ to either its echelon or its row-reduced echelon form will produce a row of zeros whenever the row is a linear combination of some (or all) of the rows above it. So if an echelon form contains $r \leq m$ nonzero rows, it follows that these r rows are linearly independent, and hence that the remaining $m - r$ rows are linearly dependent on the first r rows. The number r is called the *row rank* of matrix $\mathbf{A}$.

This means that if the r nonzero rows of an echelon form $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ are regarded as n element row vectors belonging to a vector space R^n, the r vectors will *span* a subspace of R^n. Consequently, as these vectors form a *basis* for this subspace, every vector in it can be expressed as a linear combination of the form

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_r\mathbf{u}_r,$$

where the $a_1, a_2, \ldots, a_r$ are scalar constants. This subspace of R^n is called the **row space** of matrix $\mathbf{A}$.

row and column ranks and spaces

It should be remembered that the vectors forming a basis for a space are not unique, and that any basis can be transformed to any other one by means of suitable linear combinations of the vectors involved. So although the r nonzero rows of the echelon form of $\mathbf{A}$ and those of its row-reduced echelon form look different, they are equivalent, and each forms a basis for the row space of $\mathbf{A}$.

Just as there may be linear dependence between the rows of $\mathbf{A}$, so also may there be linear dependence between its columns. If s of the n columns of an $m \times n$ matrix $\mathbf{A}$ are linearly independent, the number s is called the **column rank** of matrix $\mathbf{A}$. When the s nonzero columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s$ are regarded as m element column vectors belonging to a vector space R^m, these vectors will span a subspace of R^m.

Consequently, as these vectors form a basis for this subspace, every vector in it can be expressed as a linear combination of the form

$$b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_s\mathbf{v}_s,$$

where the $b_1, b_2, \ldots, b_s$ are scalar constants. This subspace of R^m is called the **column space** of matrix **A**.

The connection between the row and column ranks of a matrix is provided by the following theorem.

THEOREM 3.7

equality of the rank of a matrix and its transpose

The equality of the row and column ranks Let **A** be any matrix. Then the row rank and column rank of **A** are equal.

Proof Let an $m \times n$ matrix **A** have row rank r. Then in its row-reduced echelon form it must contain r columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$, in each of which only the single nonzero entry 1 appears. Call these columns the *leading columns* of the row-reduced echelon form, and let them be arranged so that in the ith column $\mathbf{v}$, the entry 1 appears in the ith row.

The row-reduced echelon form of **A** will comprise the leading columns arranged in numerical order with, possibly, columns between the ith and the $(i+1)$th leading columns in which zero elements lie below the ith row but nonzero elements may occur above it. Furthermore, there may be columns to the right of column $\mathbf{v}_r$ in which zero elements lie below the rth row but nonzero elements may lie above it.

By subtracting suitable multiples of the leading columns from any columns that lie between them or to the right of $\mathbf{v}_r$, it is possible to reduce all entries in such columns to zero. Consequently, at the end of this process, the only remaining nonzero columns will be the r linearly independent leading columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$. This establishes the equality of the row and column ranks. ■

Rank

The **rank** of matrix **A**, denoted by rank (**A**), is the value common to the row and column ranks of **A**.

THEOREM 3.8

Rank of A and $\mathbf{A}^{\mathrm{T}}$ Let **A** be any matrix. Then

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^{\mathrm{T}}).$$

Proof The columns of **A** are the rows of $\mathbf{A}^{\mathrm{T}}$, so the column rank of A is the row rank of $\mathbf{A}^{\mathrm{T}}$. However, by Theorem 3.7 these two ranks are equal, so the result is proved. ■

EXAMPLE 3.22

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 & 0 & 4 & 0 \\ 2 & 1 & 7 & 0 & 10 & 1 \\ 1 & 0 & 3 & 2 & 6 & 4 \\ 1 & 0 & 3 & 0 & 4 & 0 \end{bmatrix}.$$

Then the row-reduced echelon form of **A** is **B** (**B** $\sim$ **A**)

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 3 & 0 & 4 & 0 \\ 0 & 1 & 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

showing that the number of leading columns is 3, so the row rank of **A** is 3, and hence its *rank* is 3. Three row vectors spanning a subspace of R^6, and so forming a basis for this subspace, are the three nonzero row vectors in this 4×6 matrix,

$$\mathbf{u}_1 = [1,\ 0,\ 3,\ 0,\ 4,\ 0], \quad \mathbf{u}_2 = [0,\ 1,\ 1,\ 0,\ 2,\ 1], \quad \text{and} \quad \mathbf{u}_3 = [0,\ 0,\ 0,\ 1,\ 1,\ 2].$$

The row-reduced echelon form of $\mathbf{A}^{\mathrm{T}}$ is

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

showing that the number of leading columns is 3, confirming as would be expected that the column rank of **A** (the row rank of $\mathbf{A}^{\mathrm{T}}$) is 3. The three *row* vectors of $\mathbf{A}^{\mathrm{T}}$ spanning a subspace of R^4, and so forming a basis for this subspace, are the three nonzero rows in this 6×4 matrix, namely,

$$[1,\ 0,\ 0,\ 1], \quad [0,\ 1,\ 0,\ 0], \quad \text{and} \quad [0,\ 0,\ 1,\ 0].$$

The three linearly independent *column* vectors of **A** are obtained by transposing these vectors to obtain

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$ ■

Summary

This section introduced the important algebraic concepts of the rank of a matrix, and of the row and column spaces of a matrix. The equality of the row and column ranks of a matrix was then proved. It will be seen later that the rank of a matrix plays a fundamental role when we seek a solution of a linear algebraic system of equations.

EXERCISES 3.6

In Exercises 1 through 14 find the row-reduced echelon form of the given matrix, its rank, a basis for its row space, and a basis for its column space.

1. $\begin{bmatrix} 1 & 3 & 1 & 0 & 1 & 1 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 4 & 1 & 3 \end{bmatrix}$.

2. $\begin{bmatrix} 1 & 3 & 2 & 1 \\ 2 & 0 & 2 & 1 \\ 1 & 0 & 4 & 5 \\ 0 & 1 & 2 & 4 \end{bmatrix}$.

3. $\begin{bmatrix} 3 & 0 & 2 & 6 & 0 \\ 4 & 1 & 0 & 11 & 3 \\ 2 & 0 & 2 & 4 & 0 \\ 3 & 0 & 0 & 6 & 3 \end{bmatrix}$.

4. $\begin{bmatrix} 2 & 3 & 1 & 0 & 0 & 2 & 4 \\ 1 & 2 & 1 & 0 & 4 & 1 & 2 \end{bmatrix}$.

5. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix}$.

6. $\begin{bmatrix} 3 & 2 & 4 \\ 1 & 2 & 2 \\ 8 & 8 & 12 \end{bmatrix}$.

7. $\begin{bmatrix} 1 & 3 & 4 \\ 3 & 0 & 4 \\ 2 & 3 & 1 \\ 0 & 3 & 5 \end{bmatrix}$.

8. $\begin{bmatrix} 2 & 1 & 3 & 1 \\ 1 & 1 & 0 & 3 \\ 1 & 2 & 1 & 0 \\ 3 & 3 & 4 & 1 \\ 2 & 3 & 1 & 3 \end{bmatrix}$.

9. $\begin{bmatrix} 1 & 2 & 1 & 4 & 5 & 7 \\ 2 & 1 & 0 & 1 & 2 & 1 \\ 3 & 3 & 1 & 5 & 7 & 8 \end{bmatrix}$.

10. $\begin{bmatrix} 2 & 4 & 0 & 10 & 8 \\ 0 & 2 & 1 & 3 & 1 \\ 2 & 6 & 1 & 13 & 9 \end{bmatrix}$.

11. $\begin{bmatrix} 0 & -1 & 4 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$.

12. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 2 & 5 & -1 & 0 \\ 1 & 3 & 2 & 1 \end{bmatrix}$.

13. $\begin{bmatrix} 1 & 7 & 2 & 4 \\ 0 & 0 & 5 & 7 \\ 0 & 0 & 0 & 3 \end{bmatrix}$.

14. $\begin{bmatrix} 1 & 5 & 0 & 3 \\ 2 & 1 & 1 & 1 \\ 1 & 2 & 3 & 2 \\ 3 & 3 & 4 & 3 \\ 4 & 5 & 7 & 5 \end{bmatrix}$.

3.7 The Solution of Homogeneous Systems of Linear Equations

Having now introduced the echelon and row-reduced echelon forms of an $m \times n$ matrix $\mathbf{A}$, we are in a position to discuss the nature of the solution set of the system of linear equations

homogeneous and nonhomogeneous systems of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{20}$$

which will be **nonhomogeneous** when at least one of the terms b_i on the right is nonzero, and **homogeneous** when $b_1 = b_2 = \cdots = b_m = 0$. In this section we will only consider homogeneous systems.

Rather than working with the full system of homogeneous equations corresponding to $b_i = 0$, $i = 1, 2, \ldots, m$ in (20), it is more convenient to work with its coefficient matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \tag{21}$$

which contains all the information about the system. The coefficients in the first column of $\mathbf{A}$ are multipliers of x_1, those in the second column are multipliers of $x_2, \ldots,$ and those in the nth column are multipliers of x_n.

Denote by $\mathbf{A}_{\mathrm{E}}$ either the echelon or the row-reduced echelon form of the coefficient matrix $\mathbf{A}$. Then, as elementary row operations performed on a coefficient matrix are equivalent in all respects to performing the same operations on the corresponding full system of equations, the solution set of the matrix equation

$$\mathbf{Ax} = \mathbf{0} \tag{22}$$

will be the same as the solution set of an echelon form of the homogeneous equations

$$\mathbf{A}_{\mathrm{E}}\mathbf{x} = \mathbf{0}. \tag{23}$$

trivial solution

It is obvious that $\mathbf{x} = \mathbf{0}$, corresponding to $\mathbf{x} = [0, 0, \ldots, 0]^{\mathrm{T}}$, is always a solution of (22) and, of course of (23), and it is called the **trivial solution** of the homogeneous system of equations. To discover when nontrivial solutions exist it is necessary to work with the equivalent echelon form of the equations given in (23).

If $\operatorname{rank}(\mathbf{A}) = r$, the first r rows of $\mathbf{A}_{\mathrm{E}}$ will be nonzero rows, and the last $m - r$ rows will be zero rows. As there are m rows in $\mathbf{A}$, we must consider the three separate cases (a) $m < n$, (b) $m = n$, and (c) $m > n$.

Case (a): $m < n$. In this case there are more variables than equations. As $\operatorname{rank}(\mathbf{A}) = r$, and there are m equations, it follows that $r = \operatorname{rank}(\mathbf{A}) \le m$. The system in (22) will thus contain only r linearly independent equations corresponding to the first r rows of $\mathbf{A}_{\mathrm{E}}$. So working with system (23), we see that r of the variables $x_1, x_2, \ldots, x_n$ will be determined in terms of the remaining $m - r$ variables regarded as parameters (see Example 3.23).

Case (b): $m = n$. In this case the number of variables equals the number of equations. If $\operatorname{rank}(\mathbf{A}) = r < n$ we have the same situation as in Case (a), and the variables $x_1, x_2, \ldots, x_n$ will be determined by the system of equations in (23) in terms of the remaining $m - r$ variables regarded as parameters. However, if $r = n$, only the trivial solution $\mathbf{x} = \mathbf{0}$ is possible, because in this case $\mathbf{A}_{\mathrm{E}}$ becomes the unit matrix $\mathbf{I}_n$, from which it follows directly that $\mathbf{x} = \mathbf{0}$.

Case (c): $m > n$. In this case the number of equations exceeds the number of variables and $r = \operatorname{rank}(\mathbf{A}) \le n$. This is essentially the same situation as in Case (b), because if $r = \operatorname{rank}(\mathbf{A}) < n$, the variables $x_1, x_2, \ldots, x_n$ will be determined by the system of equations in (22) in terms of the remaining $m - r$ variables regarded as parameters, while if $\operatorname{rank}(\mathbf{A}) = n$ only the trivial solution $\mathbf{x} = \mathbf{0}$ is possible.

The practical determination of solution sets to homogeneous systems of linear equations is illustrated in the next example.

EXAMPLE 3.23

Find the solution sets of the homogeneous systems of linear equations with coefficient matrices given by:

$$\text{(a)}\ \mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 7 & 0 \\ 3 & 6 & 4 & 24 & 3 \\ 1 & 4 & 4 & 12 & 3 \end{bmatrix}, \quad \text{(b)}\ \mathbf{A} = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \quad \text{(c)}\ \mathbf{A} = \begin{bmatrix} 2 & 3 & 6 & 1 \\ 1 & 4 & 2 & 2 \\ 4 & 11 & 10 & 5 \\ 1 & 0 & 1 & 1 \end{bmatrix},$$

$$\text{(d)}\ \mathbf{A} = \begin{bmatrix} 1 & 4 & 1 & 2 \\ 1 & 3 & 0 & 1 \\ 2 & 1 & 1 & 1 \\ 4 & 9 & 3 & 5 \\ 5 & 5 & 2 & 3 \end{bmatrix}, \quad \text{(e)}\ \mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 1 & 4 & 3 \\ 0 & 1 & 3 & 0 & 1 & 5 \\ 3 & 1 & 2 & 3 & 1 & 4 \end{bmatrix}$$

Solution

(a) The row-reduced echelon form of the matrix is

$$\mathbf{A}_{\mathrm{E}} = \begin{bmatrix} 1 & 0 & 0 & 8 & 3 \\ 0 & 1 & 0 & -2 & -3 \\ 0 & 0 & 1 & 3 & 3 \end{bmatrix},$$

showing that rank(**A**) $= 3$. This corresponds to the following three equations between the five variables x_1, x_2, x_3, x_4, and x_5:

$$x_1 + 8x_4 + 3x_5 = 0, \quad x_2 - 2x_4 - 3x_5 = 0, \quad \text{and} \quad x_3 + 3x_4 + 3x_5 = 0.$$

Letting $x_4 = \alpha$ and $x_5 = \beta$ be arbitrary numbers (parameters) allows the solution set to be written

$$x_1 = -8\alpha - 3\beta, \quad x_2 = 2\alpha + 3\beta, \quad x_3 = -3\alpha - 3\beta, \quad x_4 = \alpha, \quad x_5 = \beta.$$

(b) The row-reduced echelon form of the matrix is

$$\mathbf{A}_{\mathrm{E}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

showing that rank(**A**) $= 3$. This corresponds to the trivial solution $x_1 = x_2 = x_3 = 0$.

(c) The row-reduced echelon form of the matrix is

$$\mathbf{A}_{\mathrm{E}} = \begin{bmatrix} 1 & 0 & 0 & 20/13 \\ 0 & 1 & 0 & 5/13 \\ 0 & 0 & 1 & -7/13 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

showing that rank(**A**) $= 3$. This corresponds to the solution set $x_1 + (20/13)x_4 = 0$, $x_2 + (5/13)x_4 = 0$, and $x_3 - (7/13)x_4 = 0$. Setting $x_4 = k$, an arbitrary number (a parameter), shows the solution set to be given by

$$x_1 = -(20/13)k, \quad x_2 = -(5/13)k, \quad x_3 = (7/13)k, \quad \text{and} \quad x_4 = k.$$

(d) The row-reduced echelon form of the matrix is

$$\mathbf{A}_{\mathrm{E}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & 2/3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

showing that rank(**A**) $= 3$. This corresponds to the following three equations for the four variables x_1, x_2, x_3, and x_4:

$$x_1 = 0, \quad x_2 + (1/3)x_4 = 0, \quad \text{and} \quad x_3 + (2/3)x_4 = 0.$$

Setting $x_4 = k$, an arbitrary number (a parameter), shows the solution set to be given by

$$x_1 = 0, \quad x_2 = -k/3 = 0, \quad x_3 = -2k/3, \quad \text{and} \quad x_4 = k.$$

(e) The row-reduced echelon form of the matrix is

$$\mathbf{A}_{\mathrm{E}} = \begin{bmatrix} 1 & 0 & 0 & 1 & -1/4 & 1/2 \\ 0 & 1 & 0 & 0 & 13/4 & -5/2 \\ 0 & 0 & 1 & 0 & -3/4 & 5/2 \end{bmatrix},$$

showing that rank(**A**) $= 3$. This corresponds to the following three equations for the six variables x_1 to x_6:

$$x_1 + x_4 - (1/4)x_5 + (1/2)x_6 = 0, \quad x_2 + (13/4)x_5 - (5/2)x_6 = 0$$
$$x_3 - (3/4)x_5 + (5/2)x_6 = 0.$$

Setting $x_4 = \alpha$, $x_5 = \beta$, and $x_6 = \gamma$, where α, β, and γ are arbitrary numbers (parameters), shows the solution set to be given by

$$x_1 = -\alpha + (1/4)\beta - (1/2)\gamma, \quad x_2 = -(13/4)\beta + (5/2)\gamma, \quad x_3 = (3/4)\beta - (5/2)\gamma$$
$$x_4 = \alpha, \quad x_5 = \beta, \quad \text{and} \quad x_6 = \gamma.$$

Summary

This section made use of the rank of a matrix to determine when a nontrivial solution of a linear system of homogeneous linear algebraic equations exists and, when it does, its precise form.

EXERCISES 3.7

In Exercises 1 through 10, use the given form of the matrix **A** to find the solution set of the associated homogeneous linear system of equations $\mathbf{Ax} = \mathbf{0}$.

1. $\begin{bmatrix} 1 & 3 & 2 & 1 & 1 \\ 1 & 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 1 & 3 \end{bmatrix}$.

2. $\begin{bmatrix} 1 & 2 & 0 & 1 & 1 \\ 0 & 3 & 1 & 0 & 1 \\ 2 & 0 & 2 & 0 & 1 \\ 1 & 0 & 3 & 1 & 1 \end{bmatrix}$.

3. $\begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 1 & 3 \\ 1 & 4 & 1 & 3 \\ 2 & 6 & 5 & 4 \end{bmatrix}$.

4. $\begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 0 & 3 & 5 & 1 \\ 1 & 0 & 1 & 5 \end{bmatrix}$.

5. $\begin{bmatrix} 1 & 3 & 4 \\ 2 & 1 & 3 \\ 1 & 0 & 2 \\ 3 & 1 & 1 \\ 2 & 3 & 1 \end{bmatrix}$.

6. $\begin{bmatrix} 2 & 1 & 1 & 3 \\ 1 & 2 & 3 & 0 \\ 0 & 1 & 4 & 2 \\ 1 & 3 & 1 & 2 \\ 0 & 4 & 1 & 1 \end{bmatrix}$.

7. $\begin{bmatrix} 1 & 5 & 2 & 2 & 1 & 3 & 2 \\ 0 & 1 & 4 & 1 & 0 & 1 & 1 \\ 1 & 2 & 1 & 0 & 0 & 2 & 0 \\ 2 & 3 & 0 & 1 & 1 & 0 & 2 \end{bmatrix}$.

8. $\begin{bmatrix} 1 & 4 & 1 & 0 \\ 2 & 1 & 3 & 1 \\ 5 & 6 & 7 & 2 \\ 2 & 1 & 0 & 1 \end{bmatrix}$.

9. $\begin{bmatrix} 1 & 1 & 5 & 0 & 0 & 1 \\ 2 & 3 & 1 & 2 & 1 & 3 \\ 0 & 1 & 0 & 1 & 3 & 0 \end{bmatrix}$.

10. $\begin{bmatrix} 1 & 3 & 2 & 1 & 1 \\ 2 & 5 & 1 & 0 & 2 \\ 0 & 1 & 2 & 0 & 3 \\ 1 & 0 & 3 & 1 & 2 \end{bmatrix}$.

3.8 The Solution of Nonhomogeneous Systems of Linear Equations

We now turn our attention to the solution of the nonhomogeneous system of equations in (20) that may be written in the matrix form

$$\mathbf{Ax} = \mathbf{b}, \tag{24}$$

where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an $m \times 1$ nonzero column vector. In many respects the arguments we now use parallel the ones used when seeking the form of the solution set for a homogeneous system, but there are important differences. This time, rather than working with the matrix $\mathbf{A}$, we must work with the augmented matrix $(\mathbf{A}, \mathbf{b})$ and use elementary row operations to transform it into either an echelon or a row-reduced echelon form that will be denoted by $(\mathbf{A}, \mathbf{b})_{\mathrm{E}}$. When this is done, system (24) and the echelon form corresponding to $(\mathbf{A}, \mathbf{b})_{\mathrm{E}}$ will, of course, each have the same solution set.

It is important to recognize that $\mathrm{rank}(\mathbf{A})$ is not necessarily equal to rank $(\mathbf{A}, \mathbf{b})_{\mathrm{E}}$, so that in general $\mathrm{rank}(\mathbf{A}) \leq \mathrm{rank}((\mathbf{A}, \mathbf{b})_{\mathrm{E}})$. The significance of this observation will become clear when we seek solutions of systems like (24).

Case (a): $m < n$. In this case there are more variables than equations, and it must follow that $\text{rank}((\mathbf{A}, \mathbf{b})_E) \leq m$. If $\text{rank}(\mathbf{A}) = \text{rank}((\mathbf{A}, \mathbf{b})_E) = r$, it follows that r of the equations in (24) are linearly independent and $m - r$ are linear combinations of these r equations. This means that the first r rows of $(\mathbf{A}, \mathbf{b})_E$ are linearly independent while the last $m - r$ rows are rows of zeros. Thus, r of the variables x_1 to x_n will be determined by the equations corresponding to these r nonzero rows, in terms of the remaining $m - r$ variables as parameters. It can happen, however, that $\text{rank}(\mathbf{A}) = r < \text{rank}((\mathbf{A}, \mathbf{b})_E)$, and then the situation is different, because one or more of the rows following the rth row will have zeros in its first n entries and nonzero numbers for their last entries. When interpreted as equations, these will imply contradictions, because they will assert expressions such as $0 = c$ with $c \neq 0$ that are impossible. Thus, no solution will exist if $\text{rank}(\mathbf{A}) \neq \text{rank}((\mathbf{A}, \mathbf{b})_E)$.

Case (b): $m = n$. In this case the number of variables equals the number of equations, and it must follow that $\text{rank}((\mathbf{A}, \mathbf{b})_E) \leq n$. The situation now parallels that of Case (a), because if $\text{rank}(\mathbf{A}) = \text{rank}((\mathbf{A}, \mathbf{b})_E) = r < m$, then r of the equations in (24) will be linearly independent, while $m - r$ will be linear combinations of these r equations. So, as before, the first r rows of $(\mathbf{A}, \mathbf{b})_E$ will be linearly independent while the last $m - r$ rows will be rows of zeros. Thus, r of the variables x_1 to x_n will be determined by the equations corresponding to these r nonzero rows in terms of the remaining $m - r$ variables as parameters. In the case $r = n$, the solution will be unique, because then $\mathbf{A}_E = \mathbf{I}$. Finally, if $\text{rank}(\mathbf{A}) \neq \text{rank}((\mathbf{A}, \mathbf{b})_E)$, it follows, as in Case (a), that no solution will exist.

Case (c): $m > n$. In this case there are more equations than variables, and it must follow that $\text{rank}((\mathbf{A}, \mathbf{b})_E) \leq n$. If $\text{rank}(\mathbf{A}) = \text{rank}((\mathbf{A}, \mathbf{b})_E) = r$, it follows, as in Case (b), that r of the equations in (24) are linearly independent while $m - r$ are linear combinations of these r equations. Thus, again, the first r rows of $(\mathbf{A}, \mathbf{b})_E$ will be linearly independent while the last $m - r$ rows will be rows of zeros. Consequently, r of the variables x_1 to x_n will be determined by the equations corresponding to these r nonzero rows in terms of the remaining $m - r$ variables as parameters. If $\text{rank}(\mathbf{A}) \neq \text{rank}((\mathbf{A}, \mathbf{b})_E)$, then as before no solution will exist.

These considerations bring us to the definition of consistent and inconsistent systems of nonhomogeneous equations, with consistent systems having solutions, sometimes in terms of parameters, and inconsistent systems have no solution.

consistent and inconsistent systems

Consistent and inconsistent nonhomogeneous systems

The nonhomogeneous system $\mathbf{Ax} = \mathbf{b}$ is said to be **consistent** when it has a solution; otherwise, it is said to be **inconsistent**.

As with homogeneous systems, the practical determination of solution sets of nonhomogeneous systems of linear equations will be illustrated by means of examples.

EXAMPLE 3.24 Find the solution sets for each of the following augmented matrices $(\mathbf{A}, \mathbf{b})$, where the matrices $\mathbf{A}$ are those given in Example 3.23.

(a) $(\mathbf{A}, \mathbf{b}) = \left[\begin{array}{ccccc|c} 1 & 2 & 1 & 7 & 0 & 1 \\ 3 & 6 & 4 & 24 & 3 & 0 \\ 1 & 4 & 4 & 12 & 3 & 3 \end{array}\right]$ **(b)** $(\mathbf{A}, \mathbf{b}) = \left[\begin{array}{ccc|c} 1 & 3 & 2 & 2 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & -3 \end{array}\right]$

(c) $(\mathbf{A}, \mathbf{b}) = \left[\begin{array}{cccc|c} 2 & 3 & 6 & 1 & 2 \\ 1 & 4 & 2 & 2 & 3 \\ 4 & 11 & 10 & 5 & 1 \\ 1 & 0 & 1 & 1 & 2 \end{array}\right]$ **(d)** $(\mathbf{A}, \mathbf{b}) = \left[\begin{array}{cccc|c} 1 & 4 & 1 & 2 & 2 \\ 1 & 3 & 0 & 1 & 0 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 9 & 3 & 5 & 7 \\ 5 & 5 & 2 & 3 & 0 \end{array}\right]$

(e) $(\mathbf{A}, \mathbf{b}) = \left[\begin{array}{cccccc|c} 1 & 2 & 3 & 1 & 4 & 3 & -2 \\ 0 & 1 & 3 & 0 & 1 & 5 & 0 \\ 3 & 1 & 2 & 3 & 1 & 4 & 1 \end{array}\right].$

Solution

(a) In this case,

$$(\mathbf{A}, \mathbf{b})_{\mathrm{E}} = \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 8 & 3 & -7 \\ 0 & 1 & 0 & -2 & -3 & 11/2 \\ 0 & 0 & 1 & 3 & 3 & -3 \end{array}\right].$$

As $\operatorname{rank}(\mathbf{A}, \mathbf{b})_{\mathrm{E}} = 3$, and the rank of matrix $\mathbf{A}$ is the rank of the matrix formed by deleting the last column of $(\mathbf{A}, \mathbf{b})_{\mathrm{E}}$, it follows that $\operatorname{rank}(\mathbf{A}) = 3$. So $\operatorname{rank}(\mathbf{A}, \mathbf{b})_{\mathrm{E}} = \operatorname{rank}(A)$, showing the equations to be consistent, so they have a solution.

If we remember that the first column contains the coefficients of x_1, the second column the coefficients of $x_2, \ldots,$ and the fifth column the coefficients of x_5, while the last column contains the nonhomogeneous terms, we can see that the matrix $(\mathbf{A}, \mathbf{b})_{\mathrm{E}}$ is equivalent to the three equations

$$x_1 + 8x_4 + 3x_5 = -7, \quad x_2 - 2x_4 - 3x_5 = 11/2, \quad x_3 + 3x_4 + 3x_5 = -3.$$

So, if we set $x_4 = \alpha$ and $x_5 = \beta$, with α and β arbitrary numbers (parameters), the solution set becomes

$$x_1 = -8\alpha - 3\beta - 7, \quad x_2 = 2\alpha + 3\beta + 11/2, \quad x_3 = -3\alpha - 3\beta - 3,$$
$$x_4 = \alpha \quad \text{and} \quad x_5 = \beta.$$

(b) In this case,

$$(\mathbf{A}, \mathbf{b})_{\mathrm{E}} = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 9 \\ 0 & 1 & 0 & -17 \\ 0 & 0 & 1 & 22 \end{array}\right].$$

Here $\mathbf{A}$ is a 3×3 matrix and $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}((\mathbf{A}, \mathbf{b})_{\mathrm{E}}) = 3$, so the equations are consistent and the solution is unique. The solution set is seen to be

$$x_1 = 9, \quad x_2 = -17, \quad \text{and} \quad x_3 = 22.$$

(c) In this case,

$$(\mathbf{A}, \mathbf{b})_{\mathrm{E}} = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 20/13 & 0 \\ 0 & 1 & 0 & 5/13 & 0 \\ 0 & 0 & 0 & -7/13 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This system has no solution because the equations are inconsistent. This follows from the fact that $\mathrm{rank}(\mathbf{A}) = 3$, as can be seen from the first four columns, while the five columns show that $\mathrm{rank}((\mathbf{A}, \mathbf{b})_{\mathrm{E}}) = 4$, so that $\mathrm{rank}(\mathbf{A}) \neq \mathrm{rank}((\mathbf{A}, \mathbf{b})_{\mathrm{E}})$. The inconsistency can be seen from the contradiction contained in the last row, which asserts that $0 = 1$.

(d) In this case

$$(\mathbf{A}, \mathbf{b})_{\mathrm{E}} = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/3 & 0 \\ 0 & 0 & 1 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This system also has no solution because the equations are inconsistent. This follows from the fact that $\mathrm{rank}(\mathbf{A}) = 3$ and $\mathrm{rank}((\mathbf{A}, \mathbf{b})_{\mathrm{E}}) = 4$, so that $\mathrm{rank}(\mathbf{A}) \neq \mathrm{rank}((\mathbf{A}, \mathbf{b})_{\mathrm{E}})$. The inconsistency can again be seen from the contradiction in the last row, which again asserts that $0 = 1$.

(e) In this case

$$(\mathbf{A}, \mathbf{b})_{\mathrm{E}} = \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & -1/4 & 1/2 & 5/8 \\ 0 & 1 & 0 & 0 & 13/4 & -5/2 & -21/8 \\ 0 & 0 & 1 & 0 & -3/4 & 5/2 & 7/8 \end{array}\right],$$

showing that $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}((\mathbf{A}, \mathbf{b})_{\mathrm{E}}) = 3$, so the equations are consistent.

Reasoning as in (a) and setting $x_4 = \alpha$, $x_5 = \beta$, and $x_6 = \gamma$, with α, β, and γ arbitrary numbers (parameters), shows the solution set to be given by

$$x_1 = -\alpha + (1/4)\beta - (1/2)\gamma + 5/8, \quad x_2 = -(13/4)\beta + (5/2)\gamma - 21/8,$$
$$x_3 = (3/4)\beta - (5/2)\gamma + 7/8, \quad x_4 = \alpha, \quad x_5 = \beta, \quad x_6 = \gamma.$$

■

general solution of a nonhomogeneous system

A comparison of the corresponding solution sets in Examples 3.23 and 3.24 shows that whenever the nonhomogeneous system has a solution, it comprises the sum of the solution set of the corresponding homogeneous system, containing arbitrary parameters, and numerical constants contributed by the nonhomogeneous terms. This is no coincidence, because it is a fundamental property of nonhomogeneous linear systems of equations. The combination of solutions comprising the sum of a solution of the homogeneous system $\mathbf{Ax} = \mathbf{0}$ containing arbitrary constants, and a particular fixed solution of the nonhomogeneous system $\mathbf{Ax} = \mathbf{b}$ that is free from arbitrary constants, is called the **general solution** of a nonhomogeneous system. The result is important, so it will be recorded as a theorem.

THEOREM 3.9 **General solution of a nonhomogeneous system** The nonhomogeneous system of equations

$$\mathbf{Ax} = \mathbf{b}$$

for which $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}\,((\mathbf{A}, \mathbf{b})_{\mathrm{E}})$ has a general solution of the form

$$\mathbf{x} = \mathbf{x}_{\mathrm{H}} + \mathbf{x}_{\mathrm{P}},$$

where $\mathbf{x}_{\mathrm{H}}$ is the general solution of the associated homogeneous system $\mathbf{Ax}_{\mathrm{H}} = \mathbf{0}$ and $\mathbf{x}_{\mathrm{P}}$ is a particular (fixed) solution of the nonhomogeneous system $\mathbf{Ax}_{\mathrm{P}} = \mathbf{b}$.

Proof Let $\mathbf{x}$ be any solution of the nonhomogeneous system $\mathbf{Ax} = \mathbf{b}$, and let $\mathbf{x}_{\mathrm{P}}$ be a solution of the nonhomogeneous system $\mathbf{Ax}_{\mathrm{P}} = \mathbf{b}$ that contains no arbitrary constants (a *fixed* solution). Then, as the equations are linear,

$$\mathbf{A}(\mathbf{x} - \mathbf{x}_{\mathbf{P}}) = \mathbf{Ax} - \mathbf{Ax}_{\mathbf{P}} = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

showing that the difference $\mathbf{x}_{\mathbf{D}} = \mathbf{x} - \mathbf{x}_{\mathbf{P}}$ is itself a solution of the homogeneous system. Consequently, all solutions of the nonhomogeneous system are contained in the solution set of the homogeneous system to which $\mathbf{x}_{\mathbf{D}}$ belongs, and the theorem is proved. ■

Summary

This section used the rank of a matrix to determine when a solution of a linear system of nonhomogeneous equations exists and to determine its precise form. If the ranks of a matrix and an augmented matrix are equal, it was shown that a solution exists, furthermore, if there are n equations and the rank $r < n$, then r unknowns can be expressed in terms of arbitrary values assigned to the remaining $n - r$ unknowns. The system was shown to have a unique solution when $r = n$, and no solution if the ranks of the matrix and the augmented matrix are different.

EXERCISES 3.8

In Exercises 1 through 10 write down a system of equations with an appropriate number of unknowns $x_1, x_2, \ldots$ corresponding to the augmented matrix. Find the solution set when the equations are consistent, and state when the equations are inconsistent.

1. $\left[\begin{array}{rrrr|r} 1 & -2 & 1 & 3 & 11 \\ 0 & 3 & -2 & 1 & 11 \\ 2 & 1 & 0 & 4 & 23 \\ 3 & 2 & -1 & 2 & 21 \\ 1 & -1 & 3 & 2 & 4 \end{array}\right].$

2. $\left[\begin{array}{rrrr|r} 2 & 1 & 3 & 1 & 1 \\ 0 & 1 & 4 & 1 & 1 \\ 3 & 0 & 0 & 2 & 1 \end{array}\right].$

3. $\left[\begin{array}{rrrr|r} 1 & 3 & 1 & 1 & 0 \\ 1 & 1 & 3 & 2 & 1 \\ 1 & 1 & 0 & 3 & 1 \\ 2 & 0 & 2 & 1 & 0 \end{array}\right].$

4. $\left[\begin{array}{rrrr|r} 1 & 4 & 2 & 3 & 4 \\ 2 & 0 & 3 & 1 & 2 \\ 5 & 4 & 8 & 5 & 8 \end{array}\right].$

5. $\left[\begin{array}{rrrr|r} 1 & -1 & 2 & -1 & -4 \\ 2 & 3 & 1 & 2 & 12 \\ 1 & 2 & -2 & 3 & 15 \\ 3 & 1 & -1 & 1 & 11 \\ 1 & 1 & -1 & 2 & 3 \end{array}\right].$

6. $\left[\begin{array}{rrr|r} 1 & 2 & 3 & 1 \\ 2 & 1 & 1 & 3 \\ 0 & 2 & 1 & 3 \\ 2 & 6 & 7 & 5 \\ 1 & -2 & 1 & 0 \end{array}\right].$

7. $\left[\begin{array}{rrrr|r} 1 & 2 & 3 & 0 & 1 \\ 0 & 1 & 0 & 2 & 1 \\ 2 & 1 & 3 & 1 & 0 \\ 1 & 4 & 1 & 5 & 2 \end{array}\right].$

8. $\left[\begin{array}{rrrrr|r} 2 & 1 & 0 & 0 & 3 & 1 \\ 1 & 2 & 1 & 1 & 3 & 0 \\ 0 & 1 & 2 & 5 & 1 & 2 \end{array}\right].$

9. $\left[\begin{array}{rrr|r} 1 & 2 & 1 & 4 \\ 1 & 1 & 2 & 0 \\ 2 & 1 & 1 & 4 \\ 0 & 3 & 5 & 1 \end{array}\right].$

10. $\left[\begin{array}{rrrrr|r} 1 & 3 & 1 & 1 & 2 & 1 \\ 1 & -2 & 1 & 3 & 1 & 0 \\ 2 & 0 & 1 & 0 & 3 & 0 \end{array}\right].$

3.9 The Inverse Matrix

multiplicative inverse matrix

The operation of division is not defined for matrices. However, we will see that $n \times n$ matrices $\mathbf{A}$ for which $\det\mathbf{A} \neq 0$ have associated with them an $n \times n$ matrix $\mathbf{B}$, called its **multiplicative inverse**, with the property that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

The purpose of this section will be to develop ways of finding the multiplicative inverse of a matrix, which for simplicity is usually called the **inverse** matrix, but first we give a formal definition of the inverse of a matrix.

The inverse of a matrix

Let $\mathbf{A}$ and $\mathbf{B}$ be two $n \times n$ matrices. Then matrix $\mathbf{A}$ is said to be invertible and to have an associated **inverse** matrix $\mathbf{B}$ if

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

Interchanging the order of $\mathbf{A}$ and $\mathbf{B}$ in this definition shows that if $\mathbf{B}$ is the inverse of $\mathbf{A}$, then $\mathbf{A}$ must be the inverse of $\mathbf{B}$.

To see that not all $n \times n$ matrices have inverses, it will be sufficient to try to find a matrix $\mathbf{B}$ such that the product $\mathbf{AB} = \mathbf{I}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The product $\mathbf{AB}$ is

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a + 2c & b + 2d \\ a + 2c & b + 2d \end{bmatrix},$$

so if this product is to equal the 2×2 unit matrix $\mathbf{I}$, it is necessary that

$$\begin{bmatrix} a + 2c & b + 2d \\ a + 2c & b + 2d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Equating corresponding elements in the first columns shows that this can only hold if $a + 2c = 1$ and $a + 2c = 0$, while equating corresponding elements in the second columns shows that $b + 2d = 0$ and $b + 2d = 1$, which is impossible, so matrix $\mathbf{A}$ has no inverse. In this case $\det\mathbf{A} = 0$, and we will see later why the nonvanishing of $\det\mathbf{A}$ is necessary if $\mathbf{A}$ is to have an inverse.

singular and nonsingular $n \times n$ matrices

Nonsingular and singular matrices

An $n \times n$ matrix is said to be **nonsingular** when its inverse exists, and to be **singular** when it has no inverse.

EXAMPLE 3.25

We have already seen that the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix},$$

for which det$\mathbf{A} = 0$, has no inverse and so is *singular*. However, in the case of matrix $\mathbf{A}$ that follows, a simple matrix multiplication confirms that it has associated with it an inverse $\mathbf{B}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 1 & -2 \\ 1 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix},$$

because $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. Furthermore, det$\mathbf{A} \neq 0$, so $\mathbf{A}$ is *nonsingular*, as is $\mathbf{B}$, and each is the inverse of the other. ■

Before proceeding further it is necessary to establish that, when it exists, the inverse matrix is unique.

THEOREM 3.10

Uniqueness of the inverse matrix A nonsingular matrix $\mathbf{A}$ has a unique inverse.

Proof Suppose, if possible, that the nonsingular $n \times n$ matrix $\mathbf{A}$ has the two different inverses $\mathbf{B}$ and $\mathbf{C}$. Then as $\mathbf{AC} = \mathbf{I}$, we have

$$\mathbf{B} = \mathbf{BI} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{IC} = \mathbf{C},$$

showing that $\mathbf{B} = \mathbf{C}$, so the inverse matrix is unique. ■

It is convenient to denote the inverse of a nonsingular $n \times n$ matrix $\mathbf{A}$ by the symbol $\mathbf{A}^{-1}$. This is suggested by the exponentation notation (raising to a power), because if for the moment we write $\mathbf{A} = \mathbf{A}^1$, then $\mathbf{AA}^{-1} = \mathbf{A}^{1}\mathbf{A}^{-1} = \mathbf{I}$, showing that exponents may be combined in the usual way, with the understanding that $\mathbf{A}^1\mathbf{A}^{-1} = \mathbf{A}^{(1-1)} = \mathbf{A}^0 = \mathbf{I}$.

THEOREM 3.11

basic properties of the inverse matrix

Basic properties of inverse matrices

(i) The unit matrix $\mathbf{I}$ is its own inverse, so $\mathbf{I} = \mathbf{I}^{-1}$.
(ii) If $\mathbf{A}$ is nonsingular, so also is $\mathbf{A}^{-1}$, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
(iii) If $\mathbf{A}$ is nonsingular, so also is $\mathbf{A}^{\mathrm{T}}$, and $(\mathbf{A}^{-1})^{\mathrm{T}} = (\mathbf{A}^{\mathrm{T}})^{-1}$.
(iv) If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular $n \times n$ matrices, so is $\mathbf{AB}$, and

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

(v) If $\mathbf{A}$ is nonsingular, then $(\mathbf{A}^{-1})^m = (\mathbf{A}^m)^{-1}$ for $m = 1, 2, \ldots$.

Proof We prove only (i) and (iv), and leave the proofs of (ii), (iii), and (v) as exercises. The proof of (i) is almost immediate, because $\mathbf{I}^2 = \mathbf{I}$, showing that $\mathbf{I} = \mathbf{I}^{-1}$. To prove (iv) we premultiply $\mathbf{B}^{-1}\mathbf{A}^{-1}$ by $\mathbf{AB}$ to obtain

$$\mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I},$$

which shows that $(\mathbf{AB})^{-1}$ is $\mathbf{B}^{-1}\mathbf{A}^{-1}$, so the proof is complete. ■

A simple method of finding the inverse of an $n \times n$ matrix is by means of elementary row operations, but to justify the method we first need the following theorem.

THEOREM 3.12

Elementary row operation matrices are nonsingular Every $n \times n$ matrix $\mathbf{E}$ that represents an elementary row operation is nonsingular.

Proof Every $n \times n$ matrix $\mathbf{E}$ that represents an elementary row operation is derived from the unit matrix $\mathbf{I}$ by means of one of the three operations defined at the start of Section 3.4. So, as $\operatorname{rank}(\mathbf{I}) = n$ and $\mathbf{E}$ and $\mathbf{I}$ are row similar, it follows that $\operatorname{rank}(\mathbf{E}) = n$, and so $\mathbf{E}$ is also nonsingular. ■

finding an inverse matrix using elementary row operations

We can now describe an elementary way of finding an inverse matrix by means of elementary row transformations. Let $\mathbf{A}$ be a nonsingular $n \times n$ matrix, and let $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_m$ represent a sequence of elementary row operations of Types I, II, and III that reduces $\mathbf{A}$ to $\mathbf{I}$, so that

$$\mathbf{E}_m\mathbf{E}_{m-1}\ldots\mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}.$$

Then postmultiplying this result by $\mathbf{A}^{-1}$ gives

$$\mathbf{E}_m\mathbf{E}_{m-1}\ldots\mathbf{E}_2\mathbf{E}_1\mathbf{I} = \mathbf{A}^{-1},$$

so $\mathbf{A}^{-1}$ is given by

$$\mathbf{A}^{-1} = \mathbf{E}_m\mathbf{E}_{m-1}\cdots\mathbf{E}_2\mathbf{E}_1\mathbf{I},$$

where the product of the first m matrices on the right is nonsingular because of Theorem 3.11. Expressed in words, this result states that when a sequence of elementary row operations is used to reduce a nonsingular matrix $\mathbf{A}$ to the unit matrix $\mathbf{I}$, performing the same sequence of elementary row operations on $\mathbf{I}$, in the same order, will generate the inverse matrix $\mathbf{A}^{-1}$. If matrix $\mathbf{A}$ is singular, this will be indicated by the generation of either a complete row or a complete column of zeros before $\mathbf{I}$ is reached.

If $\mathbf{A}$ is nonsingular, it is reducible to the unit matrix $\mathbf{I}$, and clearly $\det\mathbf{A} \neq 0$. However, if $\mathbf{A}$ is singular, the attempt to reduce it to $\mathbf{I}$ will generate either a row or a column of zeros, so that then $\det\mathbf{A} = 0$. The vanishing or nonvanishing of $\det\mathbf{A}$ provides a simple and convenient test for the singularity or nonsingularity of $\mathbf{A}$ whenever n is *small*, say $n \leq 3$, because only then is it a simple matter to calculate $\det\mathbf{A}$.

The practical way in which to implement this result is not to use the matrices $\mathbf{E}_i$ to reduce $\mathbf{A}$ to $\mathbf{I}$, but to perform the operations directly on the rows of the partitioned matrix $(\mathbf{A}, \mathbf{I})$, because when $\mathbf{A}$ in the left half of the partitioned matrix has been reduced to $\mathbf{I}$, the matrix $\mathbf{I}$ in the right half will have been transformed into $\mathbf{A}^{-1}$.

EXAMPLE 3.26

Use elementary row operations to find $\mathbf{A}^{-1}$ given that

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Solution We form the augmented matrix $(\mathbf{A}, \mathbf{I})$ and proceed as described earlier.

$$(\mathbf{A}, \mathbf{I}) = \left[\begin{array}{rrr|rrr} 1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \underset{\substack{\text{add row 1} \\ \text{to row 2}}}{\sim} \left[\begin{array}{rrr|rrr} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{\substack{\text{subtract row 3} \\ \text{from row 2}}}{\sim} \left[\begin{array}{rrr|rrr} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \underset{\substack{\text{subtract row 2} \\ \text{from row 3}}}{\sim} \left[\begin{array}{rrr|rrr} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 2 \end{array}\right]$$

$$\underset{\substack{\text{subtract row 3} \\ \text{from row 1}}}{\sim} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 2 & 1 & -2 \\ 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 2 \end{array}\right].$$

The 3×3 matrix on the left of this row-equivalent partitioned matrix is now the unit matrix $\mathbf{I}$, so the required inverse matrix is the one to the right of the partition, namely,

$$\mathbf{A}^{-1} = \begin{bmatrix} 2 & 1 & -2 \\ 1 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Once $\mathbf{A}^{-1}$ has been obtained, it is always advisable to check the result by verifying that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. ■

Before proceeding further we will use elementary matrices to provide the promised proof of Theorem 3.4(viii).

the proof that det(AB) = detA detB

Proof that det(AB) = detA detB Let $\mathbf{E}_1$ be a row matrix of Type I. Then if $\mathbf{A}$ is a nonsingular matrix, $\det(\mathbf{E}_\text{I}\mathbf{A}) = -\det\mathbf{A}$, because only a row interchange is involved. However, $\det(\mathbf{E}_\text{I}) = -1$, so $\det(\mathbf{E}_\text{I}\mathbf{A}) = \det\mathbf{E}_\text{I}\det\mathbf{A}$. Similar arguments show this to be true for elementary row operation matrices of the other two types, so if $\mathbf{E}$ is an elementary row operation of any type, then

$$\det(\mathbf{E}\mathbf{A}) = \det\mathbf{E}\det\mathbf{A}.$$

If $\det\mathbf{A} \neq 0$, premultiplication by a sequence of elementary row operation matrices $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_r$ will reduce $\mathbf{A}$ to $\mathbf{I}$, so performing them on $\mathbf{I}$ in the reverse order allows us to write

$$\mathbf{A} = \mathbf{E}_1\mathbf{E}_2 \ldots \mathbf{E}_r\mathbf{I} = \mathbf{E}_1\mathbf{E}_2 \ldots \mathbf{E}_r.$$

A repetition of the result $\det(\mathbf{E}\mathbf{A}) = \det\mathbf{E}\det\mathbf{A}$ shows that

$$\det\mathbf{A} = \det\mathbf{E}_1\det\mathbf{E}_2 \ldots \det\mathbf{E}_r.$$

If $\mathbf{B}$ is conformable for multiplication with $\mathbf{A}$, using the preceding result we have

$$\begin{aligned} \det(\mathbf{A}\mathbf{B}) &= \det(\mathbf{E}_1\mathbf{E}_2 \ldots \mathbf{E}_r\mathbf{B}) \\ &= \det\mathbf{E}_1\det\mathbf{E}_2 \ldots \det\mathbf{E}_r\det\mathbf{B}, \end{aligned}$$

but

$$\det\mathbf{E}_1\det\mathbf{E}_2 \ldots \det\mathbf{E}_n = \det\mathbf{A}, \quad \text{and so} \quad \det(\mathbf{A}\mathbf{B}) = \det\mathbf{A}\det\mathbf{B}.$$

To complete the proof we must show this result remains true if $\mathbf{A}$ is singular, in which case $\det\mathbf{A} = 0$. When $\det\mathbf{A} = 0$, the attempt to reduce it to the unit matrix $\mathbf{I}$ by elementary row operation matrices will fail because at one stage it will produce a determinant in which a row will contain only zero elements. Consequently, a determinant $\det\mathbf{E}_m$, say, will be zero, which is impossible, so $\det(\mathbf{AB}) = 0$. However, if $\det\mathbf{A} = 0$, then $\det\mathbf{A}\det\mathbf{B} = 0$, so that once again $\det(\mathbf{AB}) = \det\mathbf{A}\det\mathbf{B}$, and the result is proved. ■

EXAMPLE 3.27

Use (a) elementary row operations and (b) the determinant test to show matrix $\mathbf{A}$ is singular, given that

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 4 & 3 & 1 \end{bmatrix}.$$

Solution

(a) Using elementary row operations on the augmented matrix gives

$$(\mathbf{A}, \mathbf{I}) = \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 4 & 3 & 1 & 0 & 0 & 1 \end{array}\right] \underset{\substack{\text{subtract row 1} \\ \text{from row 2}}}{\sim} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 4 & 3 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{\substack{\text{subtract 4 times} \\ \text{row 1 from row 3}}}{\sim} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 0 & -1 & 1 & -4 & 0 & 1 \end{array}\right]$$

$$\underset{\substack{\text{subtract row 2} \\ \text{from row 3}}}{\sim} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -3 & -1 & 1 \end{array}\right].$$

The reduction is terminated at this stage by the appearance of a row of zeros on the matrix to the left of the partition, showing that $\mathbf{A}$ cannot be reduced to $\mathbf{I}$, and hence that $\mathbf{A}$ is singular.

(b) Applying the determinant test to $\mathbf{A}$, we find that $\det\mathbf{A} = 0$, showing that $\mathbf{A}$ is singular. Although in this case this is by far the quickest way to establish the singularity of $\mathbf{A}$, this would not have been so had the order of $\det\mathbf{A}$ been much greater than 3. This is because when $n > 3$, the effort involved in performing the elementary row operations in an attempt to reduce $\mathbf{A}$ to $\mathbf{I}$ is considerably less than the effort involved when calculating $\det\mathbf{A}$. ■

The following very different way of finding the inverse of an $n \times n$ matrix $\mathbf{A}$ is mainly of theoretical importance, though it is a practical method when n is small. The method is based on the properties of the sum of products of elements and cofactors of a determinant.

Let $\mathbf{A} = [a_{ij}]$ be an $n \times n$ matrix, $\mathbf{C} = [C_{ij}]$ be the associated $n \times n$ matrix of cofactors and form the matrix product

$$\mathbf{AC}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}.$$

If we write $\mathbf{B} = \mathbf{A}\mathbf{C}^{\mathrm{T}}$, with $\mathbf{B} = [b_{ij}]$, it follows from the rule for matrix multiplication that

$$b_{ij} = a_{i1}C_{1j} + a_{i2}C_{2j} + \cdots + a_{in}C_{nj}.$$

Thus, b_{ij} is seen to be the sum of the product of the elements of the ith row of $\mathbf{A}$ and the corresponding cofactors of the elements of the jth row of $\mathbf{A}$. It then follows from the Laplace expansion theorem for determinants that

$$b_{ij} = \det\mathbf{A}, \quad \text{for } i = j = 1, 2, \ldots, n$$

and

$$b_{ij} = 0, \quad \text{for } i \neq j.$$

Using these results in the matrix product, we find that

$$\begin{aligned}\mathbf{A}\mathbf{C}^{\mathrm{T}} &= \begin{bmatrix} \det\mathbf{A} & 0 & 0 & \cdots & 0 \\ 0 & \det\mathbf{A} & 0 & & \\ 0 & 0 & \det\mathbf{A} & \cdots & 0 \\ & \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & \det\mathbf{A} \end{bmatrix} \\ &= \det\mathbf{A}\,\mathbf{I}.\end{aligned}$$

Consequently, provided $\det\mathbf{A} \neq 0$, it follows that

$$(1/\det\mathbf{A})\mathbf{A}\mathbf{C}^{\mathrm{T}} = \mathbf{I}.$$

Writing this as

$$\mathbf{A}\{(1/\det\mathbf{A})\mathbf{C}^{\mathrm{T}}\} = \mathbf{I}$$

shows that

$$\mathbf{A}^{-1} = (1/\det\mathbf{A})\mathbf{C}^{\mathrm{T}}.$$

adjoint matrix

The matrix $\mathbf{C}^{\mathrm{T}}$, called the **adjoint** of $\mathbf{A}$ and written adj$\mathbf{A}$, is the *transpose* of the matrix of cofactors of $\mathbf{A}$. So the formula for the inverse of $\mathbf{A}$ becomes

$$\mathbf{A}^{-1} = (1/\det\mathbf{A})\,\mathrm{adj}\mathbf{A}. \tag{25}$$

We have arrived at the following definition and theorem.

Adjoint matrix

If $\mathbf{A}$ is an $n \times n$ matrix, and $\mathbf{C}$ is the associated matrix of cofactors, the transpose $\mathbf{C}^{\mathrm{T}}$ of the matrix of cofactors is called the adjoint of $\mathbf{A}$ and is written adj$\mathbf{A}$.

THEOREM 3.13

formal definition of an inverse matrix

The inverse matrix in terms of the adjoint of A Let $\mathbf{A}$ be a nonsingular $n \times n$ matrix. Then the inverse of $\mathbf{A}$ is given by

$$\mathbf{A}^{-1} = (1/\det\mathbf{A})\,\mathrm{adj}\mathbf{A}.$$

■

EXAMPLE 3.28

Use Theorem 3.13 to find $\mathbf{A}^{-1}$, given that

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Solution The matrix of cofactors

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & -1 \\ -3 & 1 & 3 \\ 3 & -1 & -5 \end{bmatrix}, \quad \text{so } \mathbf{C}^{\mathrm{T}} = \begin{bmatrix} 1 & -3 & 3 \\ -1 & 1 & -1 \\ -1 & 3 & -5 \end{bmatrix}.$$

Expanding $\det\mathbf{A}$ in terms of the elements of its first row (we already have its associated cofactors in the first row of $\mathbf{C}$) gives $\det\mathbf{A} = 1 \cdot 1 + (-1) \cdot 3 + 1 \cdot 0 = -2$, so from Theorem 3.13,

$$\mathbf{A}^{-1} = (-1/2)\mathbf{C}^{\mathrm{T}} = \begin{bmatrix} -\frac{1}{2} & \frac{3}{2} & -\frac{3}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{3}{2} & \frac{5}{2} \end{bmatrix}.$$

■

Although the result of Theorem 3.13 is of considerable theoretical importance, unless n is small, the task of evaluating the determinants involved makes it impractical for the determination of inverse matrices. In general, for large n, an inverse matrix is found by means of a computer using elementary row operations to reduce $\mathbf{A}$ to $\mathbf{I}$.

General Proof of Cramer's Rule

proof of Cramer's rule for a system of *n* equations

In conclusion, we will use Theorem 3.13 to arrive at a simple proof of Cramer's rule for the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots& \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

If we write the system as $\mathbf{Ax} = \mathbf{b}$, then, provided $\det\mathbf{A} \neq 0$, the solution can be written

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = (1/\det\mathbf{A})(\mathrm{adj}\mathbf{A})\mathbf{b} = (1/\det\mathbf{A})\mathbf{C}^{\mathrm{T}}\mathbf{b},$$

where $\mathbf{C}^{\mathrm{T}}$ is the transpose of the matrix of cofactors of $\mathbf{A}$. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)^{\mathrm{T}}$, the ith element of $\mathbf{x}$ is given by

$$x_i = (1/\det\mathbf{A})(C_{1i}b_1 + C_{2i}b_2 + \cdots + C_{ni}b_n) \quad \text{for } i = 1, 2, \ldots, n.$$

This is simply the expansion of $\det\mathbf{A}_i$ in terms of the elements of its ith column, where $\mathbf{A}_i$ is the matrix obtained from $\mathbf{A}$ by replacing the elements of the ith column by the elements of $\mathbf{b}$. This has established that

$$x_i = \det\mathbf{A}_i/\det\mathbf{A}, \quad \text{for } i = 1, 2, \ldots, n,$$

and the proof is complete. ■

More information about the material in Sections 3.4 to 3.9 is to be found in the appropriate chapters of references [2.1], [2.5], and [2.7] to [2.12].

GABRIEL CRAMER (1704–1752):
A Swiss mathematician who made many contributions to algebra and geometry. The result called Cramer's rule was, in fact, first formulated by Maclaurin around 1729 and published posthumously in his *Treatise on Algebra* (1748). The form of the rule attributed to Cramer appeared in his book *Traite des courbes algebraiques* (1750), which became a standard reference work during the remainder of the century. The work was so well written and so often quoted that after his death Cramer was, on occasions, considered to be the originator of the rule.

Summary

Division by matrices is not defined, but the introduction of a multiplicative inverse $\mathbf{A}^{-1}$ of a nonsingular $n \times n$ matrix $\mathbf{A}$, called the inverse of $\mathbf{A}$, enables certain operations that in some sense are similar to matrix division to be performed. This section gave the formal definition of the inverse of a matrix and established its most important algebraic properties. The inverse matrix was used to prove Cramer's rule for a general system of n nonhomogeneous linear algebraic equations when the determinant of the coefficient matrix is nonsingular.

EXERCISES 3.9

In Exercises 1 through 8, construct a suitable augmented matrix and find the inverse of the given matrix using elementary row operations.

1. $\begin{bmatrix} 1 & 3 & 7 \\ 2 & 1 & -1 \\ 2 & 1 & 5 \end{bmatrix}$.

2. $\begin{bmatrix} -4 & 1 & 0 \\ 1 & -3 & 1 \\ 2 & 1 & 4 \end{bmatrix}$.

3. $\begin{bmatrix} 1 & 1 & 3 \\ 5 & 2 & 1 \\ 1 & 6 & 2 \end{bmatrix}$.

4. $\begin{bmatrix} 2 & -6 & 1 \\ 1 & 3 & 4 \\ 0 & -2 & 1 \end{bmatrix}$.

5. $\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 0 \\ 2 & 4 & 1 \end{bmatrix}$.

6. $\begin{bmatrix} 3 & 0 & 1 \\ 1 & -1 & 1 \\ 0 & 4 & 5 \end{bmatrix}$.

7. $\begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & 0 & -3 & 4 \\ 0 & 1 & 2 & 5 \\ 2 & -1 & 2 & 2 \end{bmatrix}$.

8. $\begin{bmatrix} 0 & 1 & 2 & 3 \\ 2 & 2 & -4 & 2 \\ 1 & 3 & 0 & 1 \\ 3 & 1 & 1 & 0 \end{bmatrix}$.

9. Given that

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 1 \\ 1 & 4 & 0 \\ 2 & 1 & -3 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & -3 & 1 \\ 2 & 0 & 5 \\ 3 & 1 & 2 \end{bmatrix},$$

verify that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

10. Given that

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}, \text{ verify that } (\mathbf{A}^{-1})^{\mathrm{T}} = (\mathbf{A}^{\mathrm{T}})^{-1} \text{ and}$$

$(\mathbf{A}^{-1})^2 = (\mathbf{A}^2)^{-1}$.

In Exercises 11 through 16, use Theorem 3.13 to find the inverse of the given matrix, and check the result by showing that $\mathbf{AA}^{-1} = \mathbf{I}$.

11. $\begin{bmatrix} 2 & 4 & -5 \\ 2 & 7 & 1 \\ 1 & 3 & 4 \end{bmatrix}$.

12. $\begin{bmatrix} 3 & -7 & 8 \\ 1 & 4 & 3 \\ 0 & -5 & 1 \end{bmatrix}$.

13. $\begin{bmatrix} 9 & 2 & 1 \\ 1 & 4 & 10 \\ 3 & 1 & 2 \end{bmatrix}$.

14. $\begin{bmatrix} -3 & 2 & 6 \\ 2 & -1 & 7 \\ 5 & 4 & -2 \end{bmatrix}$.

15. $\begin{bmatrix} 2 & 0 & 1 & 2 \\ 3 & 1 & 3 & 4 \\ 1 & 0 & -2 & 3 \\ 1 & -2 & 2 & 7 \end{bmatrix}$.

16. $\begin{bmatrix} 0 & 1 & -4 & 1 \\ 3 & 7 & 5 & 2 \\ 1 & -2 & 6 & 0 \\ 0 & 1 & 3 & 1 \end{bmatrix}$.

In the following two exercises, use the determinant test to show the given matrix is singular, and then verify this by using elementary row operations applied to a suitable augmented matrix, as in Example 3.27. Compare the effort involved in each case.

17. $\begin{bmatrix} 0 & 2 & 1 & 0 \\ 1 & 1 & 3 & 0 \\ 2 & 1 & 4 & 2 \\ 4 & 3 & 10 & 2 \end{bmatrix}$.

18. $\begin{bmatrix} 1 & 3 & 0 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 5 \\ 0 & -1 & 1 & 2 \end{bmatrix}$.

3.10 Derivative of a Matrix

When the elements of matrix $\mathbf{A}$ are differentiable functions of a single variable, say t, so that $\mathbf{A} = \mathbf{A}[a_{ij}(t)]$, calculus can be performed on matrices, so it becomes necessary to define the derivative of a matrix. An illustration of the need for this was given in Section 3.2(e), where the matrix differential equation $\ddot{\mathbf{x}} + \mathbf{A}\mathbf{x} = \mathbf{0}$ was obtained as the system of second order differential equations determining the motion of a compound mass–spring system.

Derivative of a matrix

fundamental definition of $d\mathbf{A}/dt$

Let the $m \times n$ matrix $\mathbf{A}$ have elements $a_{ij}(t)$ that are differentiable functions of the variable t. Then the **first order derivative** of $\mathbf{A}$ with respect to t, written $d\mathbf{A}/dt$, is defined as

$$d\mathbf{A}/dt = [d(a_{ij})/dt],$$

and its ***n*th order derivative** with respect to t is defined recursively as

$$d^n\mathbf{A}/dt^n = d/dt[d^{n-1}\mathbf{A}/dt^{n-1}], \quad \text{for } n = 1, 2, \ldots,$$

with the convention that $d^0(a_{ij})/dt^0 = a_{ij}$, so that $d^0\mathbf{A}/dt^0 = \mathbf{A}$. The derivative of a constant matrix is the null (zero) matrix $\mathbf{0}$.

EXAMPLE 3.29

Find $d\mathbf{A}/dt$ and $d^2\mathbf{A}/dt^2$ given that

(a) $\mathbf{A} = \begin{bmatrix} t^2 & 3t & \cosh t \\ 2t+1 & e^t & \sin 2t \end{bmatrix}$, (b) $\mathbf{A} = \begin{bmatrix} te^t \\ \cos 3t \end{bmatrix}$.

Solution

(a) By definition,

$$d\mathbf{A}/dt = \begin{bmatrix} 2t & 3 & \sinh t \\ 2 & e^t & 2\cos 2t \end{bmatrix} \quad \text{and} \quad d^2\mathbf{A}/dt^2 = \begin{bmatrix} 2 & 0 & \cosh t \\ 0 & e^t & -4\sin 2t \end{bmatrix}.$$

(b) $d\mathbf{A}/dt = \begin{bmatrix} e^t + te^t \\ -3\sin 3t \end{bmatrix}$ and $d^2\mathbf{A}/dt^2 = \begin{bmatrix} 2e^t + te^t \\ -9\cos 3t \end{bmatrix}$. ■

THEOREM 3.14

derivative of a sum, a product, and an inverse matrix

Derivative of the sum of two matrices Let $\mathbf{A}(t)$ and $\mathbf{B}(t)$ be an $m \times n$ matrices, each with differentiable elements. Then

$$d/dt\{\mathbf{A} + \mathbf{B}\} = d\mathbf{A}/dt + d\mathbf{B}/dt.$$

Proof The result follows immediately from the definition of the sum of two matrices. ■

THEOREM 3.15

Derivative of a matrix product Let $\mathbf{A}(t)$ be an $m \times n$ matrix and $\mathbf{B}(t)$ be an $n \times q$ matrix, each with differentiable elements. Then, if the $m \times q$ matrix $\mathbf{C}(t) = \mathbf{A}(t)\,\mathbf{B}(t)$,

$$d\mathbf{C}/dt = \{d\mathbf{A}/dt\}\mathbf{B} + \mathbf{A}\{d\mathbf{B}/dt\}.$$

Proof It follows from the definition of the matrix product of two matrices **A** and **B** that are conformable for multiplication that $c_{rs} = a_{r1}b_{1s} + a_{r2}b_{2s} + \cdots + a_{rn}b_{ns}$, so each term in c_{rs} is a product of two differentiable functions. Differentiating c_{rs} establishes the theorem in which the order of the matrix products must be as shown. ■

THEOREM 3.16

Derivative of an inverse matrix Let $\mathbf{A}(t)$ be an $n \times n$ nonsingular matrix with differentiable elements. Then

$$d\mathbf{A}^{-1}/dt = -\mathbf{A}^{-1}\{d\mathbf{A}/dt\}\mathbf{A}^{-1}.$$

Proof As **A** is nonsingular, its inverse $\mathbf{A}^{-1}$ exists and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Differentiating the matrix product $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ gives

$$\{d\mathbf{A}/dt\}\mathbf{A}^{-1} + \mathbf{A}d\mathbf{A}^{-1}/dt = \mathbf{0}.$$

Premultiplication by $\mathbf{A}^{-1}$ followed by a rearrangement establishes the theorem. ■

EXAMPLE 3.30

Find $d\mathbf{A}^{-1}/dt$ given that

$$\mathbf{A} = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

Solution We have

$$d\mathbf{A}/dt = \begin{bmatrix} -\sin t & -\cos t \\ \cos t & -\sin t \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix},$$

so from Theorem 3.16

$$d\mathbf{A}^{-1}/dt = -\mathbf{A}^{-1}\{d\mathbf{A}/dt\}\mathbf{A}^{-1} = \begin{bmatrix} -\sin t & \cos t \\ -\cos t & -\sin t \end{bmatrix}.$$

In this case the result is easily checked by direct differentiation of $\mathbf{A}^{-1}$. ■

Applications of the derivative of a matrix are to be found in reference [2.11] and, for example, in connection with systems of ordinary differential equations in reference [3.15].

Summary

Matrices can occur with functions as their elements as, for example, when a matrix describes a rotation through an angle θ about the origin of a cartesian coordinate system O$\{x, y\}$, or when a column vector contains the unknown functions $u_1(t), u_2(t), \ldots, u_n(t)$ that form the solution set of a system of linear differential equations with independent variable t. Because of this, it is necessary to understand how to differentiate a matrix with respect to an independent variable that is present in functions forming its elements. This section addressed this matter by first defining the fundamental operation of differentiation

of a matrix, and then establishing the way in which it is to be applied to the sum and product of two matrices and to the inverse matrix.

EXERCISES 3.10

In Exercises 1 through 4, find $d\mathbf{C}/dt$ and $d^2\mathbf{C}/dt^2$.

1. $\mathbf{C} = \mathbf{A} + \mathbf{B}$, where $\mathbf{A} = \begin{bmatrix} t^3 & t & t\sin t \\ t^2 & \cos t & \sin 2t \end{bmatrix}$ and

$$\mathbf{B} = \begin{bmatrix} 1 & 2t^2 & \cosh t \\ t & 3 & \cos t \end{bmatrix}.$$

2. $\mathbf{C} = \mathbf{A} - \mathbf{B}$, where $\mathbf{A} = \begin{bmatrix} e^{2t} & 1 & \tan t \\ t & \sin t & \cos 3t \end{bmatrix}$ and

$$\mathbf{B} = \begin{bmatrix} 2 & 2t & \sinh t \\ t & t & \sin t \end{bmatrix}.$$

3. $\mathbf{C} = \mathbf{A} - 2\mathbf{B}$, where $\mathbf{A} = \begin{bmatrix} t+2 & 2t & t^3 \\ 3 & 3t & e^{2t} \end{bmatrix}$ and

$$\mathbf{B} = \begin{bmatrix} e^{2t} & t & t^3 \\ 1 & t^2 & \sinh t \end{bmatrix}.$$

4. $\mathbf{C} = \mathbf{A} + 3\mathbf{B}$, where $\mathbf{A} = \begin{bmatrix} (t+1)^2 & t & t^2 \\ 2t & 1 & \ln t \end{bmatrix}$ and

$$\mathbf{B} = \begin{bmatrix} t\sin t & 4 & t \\ t & t & \cosh t \end{bmatrix}.$$

In Exercises 5 and 6, use Theorem 3.15 to find $d\mathbf{C}/dt$, where $\mathbf{C} = \mathbf{AB}$, and check the result by direct differentiation of $\mathbf{C}$.

5. $\mathbf{A} = \begin{bmatrix} \sin t & -\cos 3t \\ \cos t & \sin t \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1+2t & 2\sin t \\ 2 & \cos t \end{bmatrix}$.

6. $\mathbf{A} = \begin{bmatrix} \cosh t & \cos t \\ \sinh t & \sin t \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \ln(2t) & t \\ t & \cos t \end{bmatrix}$.

In Exercises 7 and 8 find $d\mathbf{A}^{-1}/dt$ by means of Theorem 3.16 and then verify the result by direct differentiation of $\mathbf{A}^{-1}$.

7. $\mathbf{A} = \begin{bmatrix} \cos t & \sin t & 0 \\ -\sin t & \cos t & 0 \\ t^2 & t & 1 \end{bmatrix}$.

8. $\mathbf{A} = \begin{bmatrix} t^2 & 2t \\ -t & 3t \end{bmatrix}$.

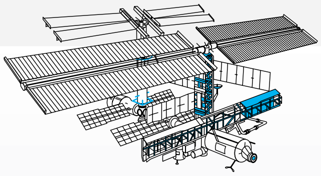
9. Find an expression for

$$d^2\{\mathbf{A}^{-1}\}/dt^2$$

in terms of $\mathbf{A}^{-1}$, $d\mathbf{A}/dt$, and $d^2\mathbf{A}/dt^2$. Apply the result to

$$\mathbf{A} = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}$$

and verify it by direct differentiation of $\mathbf{A}^{-1}$.



CHAPTER 3

TECHNOLOGY PROJECTS

Project 1

Simplification of det C When $\mathbf{C} = [c_{ij} + d_{ij}]$

The purpose of this project is to provide practice with the computer algebra of determinants and to extend the result of Theorem 3.4(vi) to the case when each element of a determinant is the sum of two numbers.

1. Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ be arbitrary 3×1 element column vectors. Then, by repeated application of Theorem 3.4(vi), extend its result to the case when $\mathbf{C} = [\mathbf{a}_1 + \mathbf{b}_1, \mathbf{a}_2 + \mathbf{b}_2, \mathbf{a}_3 + \mathbf{b}_3]$ by expressing det $\mathbf{C}$ as a sum of 3×3 determinants with columns formed from $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{b}_3$.
2. Define an arbitrary matrix $\mathbf{C}$ of the form $\mathbf{C} = [\mathbf{a}_1 + \mathbf{b}_1, \mathbf{a}_2 + \mathbf{b}_2, \mathbf{a}_3 + \mathbf{b}_3]$, and with the aid of a computer algebra determinant package find det $\mathbf{C}$ by using the result of Step 1. Confirm the result by applying the computer algebra package directly to find det $\mathbf{C}$.

Project 2

The Row-Reduced Echelon Form of a Matrix and Its Rank

The purpose of this project is to provide practice with elementary row operations performed by means of computer algebra. It involves reducing a matrix step by step, using the rules given in Section 3.5, to its row-reduced echelon form, from which its rank can then be determined by inspection.

1. Let $\mathbf{A}$ be the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 3 & 2 & 4 & 2 \\ 1 & 2 & 1 & -3 & 1 & 1 \\ -4 & 0 & 1 & 2 & 0 & 1 \\ 0 & -3 & -4 & 5 & 0 & -3 \\ 2 & 1 & -2 & -1 & 2 & -1 \end{bmatrix}.$$

   Using computer algebra, apply sequentially the steps in the rule in Section 3.5 to reduce $\mathbf{A}$ to its row-reduced echelon form, and hence find rank ($\mathbf{A}$).
2. Confirm the result obtained in Step 1 by using a computer algebra package to find directly the row-reduced echelon form of $\mathbf{A}$. Take note that in some computer algebra packages the row-reduced echelon form of a matrix $\mathbf{A}$ is called the *Gauss–Jordan form* of $\mathbf{A}$.

Project 3

A Theorem on the Rank of a Matrix Product ABC

The purpose of this project is to provide practice with matrix multiplication and the reduction of matrices to their row-reduced echelon forms using computer algebra.

1. If $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are arbitrary rectangular matrices, it can be shown that when the matrix product $\mathbf{ABC}$ exists, then

$$\text{Rank}(\mathbf{AB}) + \text{Rank}(\mathbf{BC}) \le \text{Rank}(\mathbf{B}) + \text{Rank}(\mathbf{ABC}).$$

2. Define three arbitrary rectangular matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ for which the product $\mathbf{ABC}$ is defined. Using computer algebra matrix multiplication and computer algebra row-reduction to echelon form, find the ranks of $\mathbf{AB}$, $\mathbf{BC}$, $\mathbf{B}$, and $\mathbf{ABC}$, and hence confirm the inequality in Step 1 for this particular case.

Project 4

Consistency of Augmented Coefficient Matrices, Solution by Back Substitution and Cramer's Rule

The purpose of this project is to use computer algebra to determine the consistency of two 6×7 augmented coefficient matrices. The solution for the corresponding consistent set of linear equations is then found after the reduction of its augmented coefficient matrix to row-reduced echelon form followed by back

substitution. Finally, the solution is checked using Cramer's rule, which, despite the large determinants involved, becomes feasible when computer algebra is used.

1. Use computer algebra to determine which of the augmented coefficient matrices **A** and **B** is consistent, given that

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 & 3 & 0 & 2 & 4 \\ 3 & 1 & 0 & 2 & -3 & 4 & 1 \\ 1 & 2 & 1 & 1 & 4 & 3 & 2 \\ 2 & 4 & 0 & 0 & 1 & 6 & 3 \\ 0 & 1 & 2 & 1 & -2 & 1 & 0 \\ 2 & 5 & 2 & 1 & -1 & 7 & 5 \end{bmatrix} \quad \text{and}$$

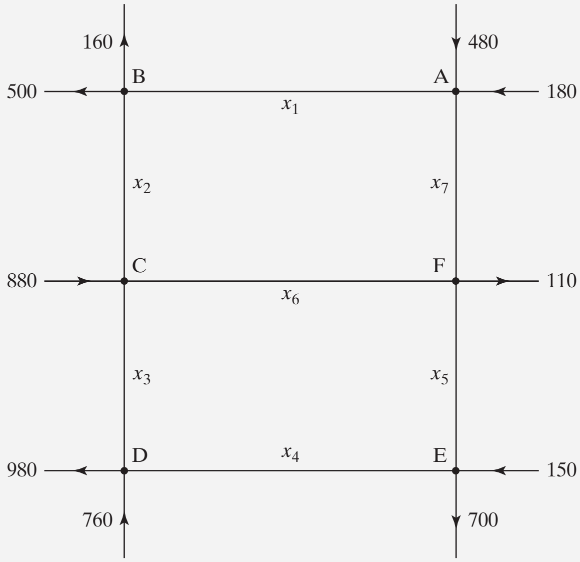
$$\mathbf{B} = \begin{bmatrix} 4 & -1 & 3 & 0 & 1 & 4 & 2 \\ 1 & 1 & -1 & -3 & 2 & 1 & -1 \\ 0 & 1 & -1 & -2 & 2 & 1 & 3 \\ 4 & 0 & 1 & -1 & 2 & 3 & -4 \\ 1 & -1 & 3 & 2 & -4 & 2 & -1 \\ 0 & 4 & 3 & -3 & 1 & 2 & 0 \end{bmatrix}.$$

2. In the case of the consistent set of equations, using the reduction of the coefficient matrix to its row-reduced echelon form, find the solution by back substitution.
3. Using computer algebra, apply Cramer's rule to the consistent set of equations to find the solution, and so confirm the result found in step 2.

Project 5

A One-Way Traffic Flow Problem

The diagram shows the pattern of one way traffic flow at six road intersections at the corners of two city blocks. The arrows show the directions of traffic flow, and the associated numbers are the traffic flow rates in vehicles per hour at peak traffic time.



By equating the flow rate of traffic into an intersection to the flow rate out of it (no parking is allowed), find equations relating the traffic flow rates $x_1, x_2, \ldots, x_7$ along each of the roads. Explain why with the given peak flow rates it is impossible to close road DE, and comment on the effect on traffic flow if road CD is closed for repairs.

Project 6

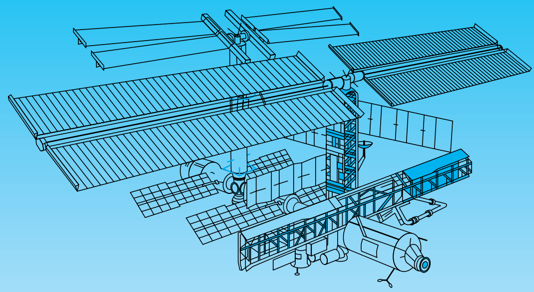
Forces in Bridge Struts

Use matrix methods to find the forces in the pin-jointed framed bridge section shown in Fig. 3.10, given that a concentrated load m acts vertically downwards at joint B.

Give a simple example of a pin-jointed framed structure that contains a redundant strut, and prove its redundancy by attempting to determine the forces acting in the strut when the structure is loaded.

This Page Intentionally Left Blank

CHAPTER 4



Eigenvalues, Eigenvectors, and Diagonalization

In engineering and physics, problems involving n linear algebraic equations in n independent variables with a constant coefficient matrix $\mathbf{A}$ often arise where a solution vector $\mathbf{x}$ is required to be proportional to $\mathbf{Ax}$. Setting the constant of proportionality equal to λ, this means that $\mathbf{x}$ must be a solution of the equation $\mathbf{Ax} = \lambda\mathbf{x}$ or, equivalently, of the equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$. The numbers λ_i for which nonzero solutions $\mathbf{x}_i$ exist are called the eigenvalues of matrix $\mathbf{A}$, and the corresponding vectors $\mathbf{x}_i$ are called the eigenvectors of $\mathbf{A}$.

Eigenvalues and eigenvectors arise, for example, when studying vibrational problems, where the eigenvalues represent fundamental frequencies of vibration and the eigenvectors characterize the corresponding fundamental modes of vibration.

They also occur in many other ways; in mechanics, for example, the eigenvalues can represent the principal stresses in a solid body, in which case the eigenvectors then describe the corresponding principal axes of stress caused by the body being subjected to external forces. Also in mechanics, the moment of inertia of a solid body about lines through its center of gravity can be represented by an ellipsoid, with the length of a line drawn from its center to the surface of the ellipsoid proportional to the moment of inertia of the body about an axis through the center of gravity of the body drawn parallel to the line. In this case the eigenvalues represent the principal moments of inertia of the body about the principal axes of inertia, that are then determined by the eigenvectors.

More precisely, if $\mathbf{A}$ is an $n \times n$ matrix, the polynomial $P_n(\lambda)$ of degree n in the scalar λ defined as $P_n(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$ is called the characteristic polynomial of $\mathbf{A}$. The roots of the equation $P_n(\lambda) = 0$ are called the eigenvalues of matrix $\mathbf{A}$, and the column vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ satisfying the matrix equation $(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{x}_i = \mathbf{0}$ are called the eigenvectors of matrix $\mathbf{A}$.

This chapter explains how eigenvalues and eigenvectors are determined and establishes important properties of eigenvectors. The eigenvectors of an $n \times n$ matrix $\mathbf{A}$ with n linearly independent eigenvectors are then used to simplify the structure of $\mathbf{A}$ by means of a process called diagonalization. An important application of diagonalization will arise later when considering the solution of linear systems of ordinary differential equations that arise from the study of mechanical, electrical, and chemical reaction problems. Diagonalization is also an important tool when working with partial differential equations, different types of which describe the temperature distribution in a metal, electromagnetic wave propagation, and diffusion processes, to name a few examples.

After a brief discussion of some special $n \times n$ matrices with complex elements, real quadratic forms are defined and the properties of eigenvectors are used to reduce a general quadratic form to a sum of squares. This is a process that finds many different applications, one of which occurs later when classifying the partial differential equations of engineering and physics in order to know the type of auxiliary conditions that must be imposed in order for them to give rise to physically meaningful solutions.

The chapter ends with the introduction of the matrix exponential $e^{\mathbf{A}}$, where $\mathbf{A}$ is a real $n \times n$ matrix, and it is shown how this enters into the solution of a linear first order matrix differential equation of the form $d\mathbf{x}/dt = \mathbf{A}\mathbf{x}$.

4.1 Characteristic Polynomial, Eigenvalues, and Eigenvectors

Throughout this chapter we will be considering the solutions of the homogeneous system of algebraic equations

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \tag{1}$$

where $\mathbf{A}[a_{ij}]$ is an $n \times n$ matrix, $\mathbf{x}$ is an n element column vector with elements $x_1, x_2, \ldots, x_n$, and λ is a scalar. For $\mathbf{A}$ given we wish to find x and λ. Introducing the $n \times n$ unit matrix by $\mathbf{I}$ allows (1) to be written

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}, \tag{2}$$

showing that $\mathbf{x}$ is a solution of a homogeneous system of equations with the coefficient matrix $\mathbf{A} - \lambda\mathbf{I}$. It was seen in Chapter 3 that nontrivial solutions $\mathbf{x}$ of (2) are only possible if one or more rows of the coefficient matrix $\mathbf{A} - \lambda\mathbf{I}$ are linearly dependent on its remaining rows. This means that nontrivial solutions $\mathbf{x}$ will exist if $\operatorname{rank}(\mathbf{A} - \lambda\mathbf{I}) < n$, but this, in turn, is equivalent to the more convenient condition $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. This is a polynomial equation for λ.

Let $P_n(\lambda)$ be the polynomial of degree n in λ defined by the determinant

$$P_n(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} - \lambda & a_{34} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} - \lambda \end{vmatrix}. \tag{3}$$

Inspection of the determinant defining $P_n(\lambda)$ shows the coefficient of λ^n is $(-1)^n$, so the polynomial is of the form

$$P_n(\lambda) = (-1)^n[\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} - \cdots + c_{n-1}\lambda + c_0]. \tag{4}$$

characteristic polynomial, equation, and eigenvalue

The polynomial $P_n(\lambda)$ is called the **characteristic polynomial** of $\mathbf{A}$ and the associated polynomial equation $P_n(\lambda) = 0$ is the **characteristic equation** of $\mathbf{A}$. As the characteristic equation of $\mathbf{A}$ is of degree n in λ, it will have n roots, some of which may be repeated. The roots of $P_n(\lambda) = 0$, or equivalently the zeros of $P_n(\lambda)$, are called the **eigenvalues** of $\mathbf{A}$ or, sometimes, the **characteristic values** of $\mathbf{A}$.

Eigenvalues (characteristic values) of A

The eigenvalues of an $n \times n$ matrix $\mathbf{A}$ are the n zeros of the polynomial $P(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$, or, equivalently, the n roots of the nth degree polynomial equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

In general, a matrix with complex coefficients will have complex eigenvalues, though even when the coefficients of $\mathbf{A}$ are all real it is still possible for complex eigenvalues to arise. This is because then the characteristic equation will have real coefficients, so if complex roots occur they must do so in complex conjugate pairs.

spectrum and spectral radius

If an eigenvalue λ^* is repeated r times, corresponding to the presence of a factor $(\lambda - \lambda^*)^r$ in the characteristic polynomial $P_n(\lambda)$, the number r is called the **algebraic multiplicity** of the eigenvalue λ^*. The set of all eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$ is called the **spectrum** of $\mathbf{A}$, and the number $R = \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\}$, equal to the largest of the moduli of the eigenvalues, is called the **spectral radius** of $\mathbf{A}$. The name comes from the fact that when the spectrum of $\mathbf{A}$ is plotted as points in the complex plane, they all lie inside or on a circle of radius R centered on the origin.

eigenvectors and eigenvalues

An **eigenvector** of an $n \times n$ matrix $\mathbf{A}$, corresponding to an eigenvalue $\lambda = \lambda_i$, is a nonzero n-element column vector $\mathbf{x}_i$ that satisfies the matrix equation

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$$

or, equivalently, that is a solution of the homogeneous system of n algebraic equations

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x}_i = \mathbf{0}. \tag{5}$$

Eigenvectors of A

The eigenvector $\mathbf{x}_i$ of the $n \times n$ matrix $\mathbf{A}$, corresponding to the eigenvalue $\lambda = \lambda_i$, is a solution of the homogeneous equation $(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x}_i = \mathbf{0}$.

It is important to recognize that because system (5) is homogeneous, the elements of an eigenvector can only be determined as multiples of one of its nonzero elements as a parameter. This means that if for some choice of the parameter $\mathbf{x}$ is an eigenvalue, then $k\mathbf{x}$ will also be an eigenvalue for any $k \neq 0$.

The next theorem is fundamental to the use of eigenvectors and shows that when an $n \times n$ matrix $\mathbf{A}$ has n distinct (different) eigenvalues, its n eigenvectors form a basis for the vector space associated with the matrix $\mathbf{A}$.

THEOREM 4.1

eigenvectors are linearly independent

Linear independence of eigenvectors The eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, corresponding to m distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, of an $n \times n$ matrix $\mathbf{A}$, are linearly independent. Furthermore, if $m = n$, the set of eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ forms a basis for the n-dimensional vector space associated with $\mathbf{A}$.

Proof The proof will be by induction, starting with two vectors, and it uses the fact that $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for $i = 1, 2, \ldots, m$.